Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Gabriel Lendel

Úlohy stochastického dynamického programování: teorie a aplikace

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce:	Ing. Karel Sladký CSc.
Studijní program:	Matematika
Studijní obor:	Pravděpodobnost, matematická statistika a ekonometrie

Praha 2011

Týmto by som rád vyjadril svoje poďakovanie vedúcemu mojej diplomovej práce Ing. Karlovi Sladkému CSc. za návrh zaujímavej témy, podnetné konzultácie, pomoc s niektorými dôkazmi a iné hodnotné rady pri spisovaní mojej práce. Ďalej chcem poďakovať Bc. Richardovi Košlabovi za pomoc s programovým vybavením práce. V neposlednom rade ďakujem svojim rodičom, ktorí mi umožnili štúdium na vysokej škole.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ..... Podpis autora

Názov práce: Úlohy stochastického dynamického programování: teorie a aplikace
Autor: Gabriel Lendel
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedúci diplomovej práce: Ing. Karel Sladký CSc.
e-mail vedúceho: sladky@utia.cas.cz

Abstrakt: V predloženej práci študujeme riadené Markovove reťazce s ohodnotením, ktoré umožňujú modelovať dynamické systémy, ktorých správanie je čiastočne náhodné a čiastočne pod kontrolou. Zaoberáme sa zostavením iteračných postupov, ktorých cielom je nájsť riadenie systému tak, aby bolo optimálne alebo skoro optimálne vzhľadom k zvolenému kritériu. Konkrétne v práci skúmame hlavne úlohu nájdenia riadenia, ktoré je optimálne vzhľadom k celkovému očakávanému diskontovanému výnosu alebo priemernému očakávanému výnosu, či pre diskrétne alebo spojité systémy. Študujeme algoritmy iterujúce riadenie (policy iteration) a aproximatívne algoritmy iterujúce hodnotu (value iteration). Pre vybrané postupy uvádzame numerickú analýzu konkrétnych problémov.

Kľúčové slová: Stochastické dynamické programovanie, riadené Markovove reťazce s ohodnotením, policy iteration, value iteration

Title: Stochastic Dynamic Programming Problems: Theory and Applications
Author: Gabriel Lendel
Department: Department of Probability and Mathematical Statistics
Supervisor: Ing. Karel Sladký CSc.
Supervisor's e-mail address: sladky@utia.cas.cz

Abstract: In the present work we study Markov decision processes which provide a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker. We study iterative procedures for finding policy that is optimal or nearly optimal with respect to the selected criteria. Specifically, we mainly examine the task of finding a policy that is optimal with respect to the total expected discounted reward or the average expected reward for discrete or continuous systems. In the work we study policy iteration algorithms and aproximative value iteration algorithms. We give numerical analysis of specific problems.

Keywords: Stochastic dynamic programming, Markov decision process, policy iteration, value iteration

Obsah

Úvod

Riadené Markovove procesy s ohodnotením, ktoré sú rozšírením Markovových reťazcov, poskytujú matematický rámec na modelovanie procesu rozhodovania v situáciách, kde je výsledok čiastočne náhodný a čiastočne pod kontrolou kontrolóra alebo automatizovaného systému. Takéto Markovove procesy sú užitočné pre štúdium širokej škály optimalizačných problémov a je možné nimi modelovať mnoho dynamických systémov v robotike, ekonómii, priemyselnej výrobe a iných odvetviach. V tejto diplomovej práci sa budeme venovať úlohám stochastického dynamického programovania, t.j. optimalizácií systémov, ktorých časový vývoj je možné popísať práve vyššie zmienenými Markovovými procesmi s ohodnotením. Táto problematika je intenzívne študovaná od 60. rokov minulého storočia, keď na koncept dynamického programovania, ktorý vyvinul Bellman, naviazal Howard, ktorý v roku 1960 odvodil policy iteration algoritmus na hľadanie optimálneho riadenia dynamického systému pracujúceho v nekonečnom časovom horizonte.

V tejto práci naviažeme na prednášku z náhodných procesov a v 1.kapitole odvodíme základné vlastnosti diskrétnych či spojitých Markovových procesov s ohodnotením. Tieto vlastnosti potom použijeme ku stavbe algoritmických postupov, ktoré hľadajú optimálne riadenie systémov vzhľadom k rôznym kritériám. Takýmto kritériom môže byť napríklad priemerný výnos, celkový diskontovaný výnos a iné. Prvým cielom je naštudovať dôkazové techniky a na základe podobností jednotlivých problémov dokázať s pomocou školiteľa sériu tvrdení, ktoré sú potrebné k zostaveniu niekoľkých algoritmov typu policy a value iteration. Ako východiskové techniky poslúžia dôkazové metódy k odvodeniu policy iteration algoritmu pre nájdenie riadenia, ktoré je optimálne vzhľadom k priemernému výnosu, tak ako sú popísané v knihe [7]. V 2.kapitole sa budeme zaoberať riadením reťazcov s diskrétnym časom, pričom na riešenie vybraných problémov zostavíme v jazyku JAVA počítačový program. Program využijeme na vyriešenie niekoľkých numerických problémov. V 3.kapitole budeme skúmať možnosť riadenia spojitých procesov, pričom algoritmické postupy na hľadanie optimálnych riadení dokážeme samostatne na základe analógie s diskrétnymi reťazcami.

Kapitola 1

Markovove reťazce s ohodnotením

1.1 Úvod

V oblasti stochastického modelovania sa ukázalo, že jednoduché modely, akými sú Markovove reťazce, sú často najužitočnejším nástrojom na analýzu praktických problémov. Teória Markovských procesov má dnes široké využitie v rôznych odvetviach vrátane biológie, techniky a operačného výskumu. V konkrétnych aplikáciách modelovanie spočíva vo vhodnej definícii stavov tak, aby mal súvisiaci reťazec Markovskú vlastnosť, t.j. že znalosť aktuálneho stavu je postačujúca k predpovedi budúceho správania procesu. V tejto kapitole zavedieme do dynamickej štruktúry reťazce aj spôsob ohodnotenia a definujeme základné pojmy teórie ohodnotených Markovských procesov.

1.2 Diskrétne reťazce

Nech je daný homogénny Markovov reťazec $\{X_n,\ n \in N_0\}$ s diskrétnym časom a konečnou množinou stavov S. Popri matici pravdepodobností prechodu $\mathbf{P} = (p_{ij})_{i,j\in S}$ uvažujme aj reálnu maticu ocenení $\mathbf{Z} = (z_{ij})_{i,j\in S}$. S prechodom zo stavu i do stavu j nech je spojený výnos alebo náklad rovný z_{ij}. Závisí od charakteru úlohy, v akom čase sa bude ocenenie realizovať. Ak nebude povedané inak, budeme predpokladať, že ku kalkulácii ocenenia dochádza v čase východiskového stavu i. Očakávaný výnos spojený s realizáciu jedného prechodu zo stavu i je potom zrejme rovný

$$c_i = \sum_{j\in S} p_{ij} z_{ij}. \tag{1.1}$$

Definujme diskontný faktor $\beta \in (0, 1]$, ktorý nám umožní zohľadniť časovú hodnotu peňazí a označme $V_i^{\beta}(n),\ i \in S$ očakávaný výnos za $n \in N_0$ období, ak bol na počiatku reťazec v stave i a ocenenie diskontujeme faktorom β. Zrejme môžeme položiť $V_i^{\beta}(0) = 0$. Diskontovať budeme do počiatku. Zrejme $\beta = 1$ znamená absenciu diskontovania.

Veta 1.1 *Pre očakávaný výnos za n období platí vzťah*

$$V_i^\beta(n) = c_i + \beta \sum_{j\in S} p_{ij} V_j^\beta(n-1),\ i \in S,\ n \in N \tag{1.2}$$

Dôkaz: Zvoľme $i \in S, n \in N$. Ak nastane jav $[X_0 = i, X_1 = i_1, \ldots, X_n = i_n]$ obdržíme výnos $z_{ii_1} + \beta z_{i_1 i_2} + \cdots + \beta^{n-1} z_{i_{n-1} i_n}$. Postupným podmieňovaním s využitím Markovskej vlastnosti odvodíme, že táto realizácia nastane s pravdepodobnosťou

$$\begin{aligned} & P(X_n = i_n, \ldots, X_1 = i_1,\ X_0 = i) = \\ = \ & P(X_{n-1} = i_{n-1}, \ldots, X_1 = i_1,\ X_0 = i)\ p_{i_{n-1} i_n} \\ = \ & P(X_{n-2} = i_{n-2}, \ldots, X_1 = i_1,\ X_0 = i)\ p_{i_{n-2} i_{n-1}}\ p_{i_{n-1} i_n} \\ = \ & p_i\ p_{ii_1}\ p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}}\ p_{i_{n-1} i_n}, \end{aligned} \tag{1.3}$$

kde sme označili $p_i = P(X_0 = i)$. Výpočtom dostaneme

$$\begin{aligned} V_i^\beta(n) \ &= \ \sum_{i_1} \cdots \sum_{i_n} (z_{ii_1} + \beta z_{i_1 i_2} + \cdots + \beta^{n-1} z_{i_{n-1} i_n}) p_i p_{ii_1} \ldots p_{i_{n-1} i_n} = \\ &= \ \sum_{i_1} p_{ii_1} \Bigg[z_{ii_1} \sum_{i_2} \cdots \sum_{i_n} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} + \\ &+ \ \beta \sum_{i_2} \cdots \sum_{i_n} (z_{i_1 i_2} + \cdots + z_{i_{n-1} i_n}) p_{i_1 i_2} \ldots p_{i_{n-1} i_n} \Bigg] \\ &= \ \sum_{i_1} p_{ii_1} z_{ii_1} + \beta \sum_{i_1} p_{ii_1} V_{i_1}^\beta(n-1) = c_i + \beta \sum_{i_1 \in S} p_{ii_1} V_{i_1}^\beta(n-1). \end{aligned} \tag{1.4}$$

Navyše vidíme, že $V_i^\beta(1) = c_i$. □

Poznámka 1.2 *Definujme stĺpcové vektory* $\mathbf{c} = \{c_i, i \in S\}$ *a* $\mathbf{V}^\beta(n) = \{V_i^\beta(n),\ i \in S\}$. *Predchádzajúce tvrdenie potom môžeme maticovo vyjadriť ako*

$$\mathbf{V}^\beta(n) = \mathbf{c} + \beta \mathbf{P} \mathbf{V}^\beta(n-1),\ n \in N.$$

Postupným dosadením do tohto rekurzívneho vzorca dostaneme

$$\mathbf{V}^\beta(n) = \mathbf{c} + \beta \mathbf{P} \mathbf{V}^\beta(n-1) = \mathbf{c} + \beta \mathbf{P}(\mathbf{c} + \beta \mathbf{P} \mathbf{V}^\beta(n-2)) = \cdots = \sum_{k=0}^{n-1} \beta^k \mathbf{P}^k \mathbf{c} \tag{1.5}$$

Poznámka 1.3 *V ďalšom texte budeme dodržiavať, že ak* $\beta = 1$ *budeme symbol* β *vynechávať, takže pri absencii diskontovania budeme písať*

$$V_i(n) = c_i + \sum_{j\in S} p_{ij} V_j(n-1),\ i \in S,\ n \in N,$$

čo môžeme maticovo vyjadriť ako

$$\mathbf{V}(n) = \mathbf{c} + \mathbf{P}\mathbf{V}(n-1),\ n \in N. \tag{1.6}$$

Postupným dosadzovaním do tohto rekurentného vzorca dostaneme

$$\mathbf{V}(n) = \sum_{k=0}^{n-1} \mathbf{P}^k \mathbf{c} \tag{1.7}$$

Naopak prítomnsť symbolu beta bude vždy znamenať diskontovanie s $\beta \in (0,1)$.

Poznámka 1.4 *Všimnime si, že rovnice* (1.3) *a* (1.4) *platia aj v prípade, keď sú uvažované pravdepodobnosti prechodu prvkami rôznych prechodových matíc. Táto skutočnosť sa významne uplatní v prípade riadených Markovských reťazcov, pretože nie je nutné uvažovať celú trajektóriu uvažovaného procesu, znalosť aktuálneho stavu poskytuje vyčerpávajúcu informáciu (hovoríme o takzvaných Markovských riadeniach).*

V praktických aplikáciách nás najčastejšie zaujíma celkový diskontovaný očakávaný výnos, ktorý dynamický systém vygeneruje alebo priemerný výnos za jedno časové obdobie v dlhodobo pracujúcom systéme. Predtým ako tieto pojmy matematicky definujeme, vyslovme základné tvrdenie, ktoré zaručuje existenciu takzvaných Cesarovských limít.

Veta 1.5 *(Existencia Cesarovských limít)*
Nech $\{X_n,\ n \in N_0\}$ *je homogénny Markovov reťazec. Ďalej nech* μ_{jj} *značí strednú dobu do prvého návratu do stavu* j *ak je východiskovým stavom práve stav* j *a* f_{ij} *pravdepodobnosť, že reťazec vôbec niekedy vstúpi do stavu* j *ak je východiskovým stavov stav* i. *Potom pre* $\forall i, j \in S$, $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}$ *vždy existuje a navyše je pre* $j \in S$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{jj}^{(k)} = \begin{cases} \frac{1}{\mu_{jj}} & \text{ak stav } j \text{ je trvalý} \\ 0 & \text{ak stav } j \text{ je prechodný} \end{cases}$$

Ďalej pre $\forall i, j \in S$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = f_{ij} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}.$$

Dôkaz: Napríklad v [7] veta 3.3.1

Definícia 1.6 Pre diskrétny Markovov reťazec s ohodnotením definujeme za predpokladu $X_0 = i,\ i \in S$ celkový očakávaný diskontovaný výnos predpisom

$$V_i^{\beta} = \lim_{n\to\infty} V_i^{\beta}(n) \tag{1.8}$$

a priemerný očakávaný výnos za časovú jednotku predpisom

$$g_i = \lim_{n\to\infty} \frac{1}{n} V_i(n). \tag{1.9}$$

Zaveďme ešte vektorové označenie $\mathbf{V}^\beta = \{V_i^\beta,\ i \in S\}$.

Podľa (1.5) máme pre diskontovaný výnos pri akomkoľvek počiatočnom stave i odhad

$$\min_{j\in S} c_j \sum_{k=0}^{n-1} \beta^k \le V_i^\beta(n) = \sum_{k=0}^{n-1} \sum_{j\in S} \beta^k p_{ij}^{(k)} c_j \le \max_{j\in S} c_j \sum_{k=0}^{n-1} \beta^k.$$

Pri $n \to \infty$ teda dostávame ohraničenie

$$(1-\beta)^{-1} \min_{j\in S} c_j \le V_i^\beta \le (1-\beta)^{-1} \max_{j\in S} c_j \tag{1.10}$$

Ďalej máme podľa (1.5) a vety 1.5 pre $\forall i \in S$

$$\begin{aligned} g_i &= \lim_{n\to\infty} \frac{1}{n} V_i(n) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j\in S} p_{ij}^{(k)} c_j \\ &= \sum_{j\in S} c_j \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} = \sum_{j\in I} c_j \frac{f_{ij}}{\mu_{jj}}, \end{aligned}$$

kde sme označili $I \subset S$ množinu trvalých stavov. Pretože je $\sum_{j\in I} \frac{f_{ij}}{\mu_{jj}} = 1$ máme pre priemerný výnos ohraničenie

$$\min_{j\in S} c_j \le g_i \le \max_{j\in S} c_j \tag{1.11}$$

Diskontovaný očakávaný výnos a priemerný očakávaný výnos sú teda pre akýkoľvek počiatočný stav konečné ohraničené hodnoty. Ďalej odvodíme explicitné vťahy pre výpočet oboch hodnôt. V prípade diskontovaného výnosu je vťah dôsledkom nasledujúceho tvrdenia.

Lemma 1.7 *Nech* $\mathbf{A}$ *je štvorcová matica taká, že* $\lim_{n\to\infty} \mathbf{A}^n = \mathbf{0}$. *Potom je matica* $\mathbf{I} - \mathbf{A}$ *regulárna a platí*

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k.$$

Dôkaz: v [3] veta B.2

Dôsledok 1.8 *Nech* $\beta \in (0,1)$. *Potom pre diskontovaný očakávaný výnos platí*

$$\mathbf{V}^\beta = (\mathbf{I} - \beta\mathbf{P})^{-1}\mathbf{c}. \tag{1.12}$$

Dôkaz: Pretože platí $\beta^k \mathbf{P}^k \to 0$, máme podľa (1.5) a lemmy 1.7

$$\mathbf{V}^\beta = \lim_{n\to\infty} \mathbf{V}^\beta(n) = \sum_{k=0}^{\infty} \beta^k \mathbf{P}^k \mathbf{c} = (\mathbf{I} - \beta\mathbf{P})^{-1}\mathbf{c}.$$

□

Vypočítať priemerný očakávaný výnos je o niečo komplikovanejšie. Ak nebude povedané inak, budeme sa ďalej vždy obmedzovať na reťazce, ktoré majú len jedinú triedu trvalých stavov, t.j. reťazce, ktoré neobsahujú 2 a viac dizjunktných uzavretých množín trvalých stavov. Ak má totiž reťazec len jedninú triedu trvalých stavov, potom k reťazcu prislúcha jednoznačne stacionárne rozdelenie $\{\Pi_j,\ j \in S\}$ a pre $\forall\ j \in S$ platí

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \Pi_j,$$

nezávisle na stave i. Pre každý východiskový stav i bude teda priemerný očakávaný výnos rovnaký, a preto budeme písať $g = g_i, i \in S$. Stacionárne rozdelenie nájdeme klasicky ako riešene systému lineárnych rovníc

$$\Pi_j = \sum_{i\in S} \Pi_i p_{ij},\ j \in S \tag{1.13}$$

$$\sum_{j\in S} \Pi_j = 1.$$

Po dosadení dostaneme, že $\forall i \in S$

$$g = \sum_{j\in S} \Pi_j c_j. \tag{1.14}$$

Nech je daná nerozložitelná matica pravdepodobností prechodu $\mathbf{P}$, t.j. všetky stavy sú trvalé. Pre ľubovolne zvolený stav $i_0 \in S$ nás ďalej môže zaujímať celkový očakávaný výnos do prvého vstupu do stavu i_0. Definujme maticu $\bar{\mathbf{P}}$ ako

$$\bar{p}_{ij} = \begin{cases} p_{ij} & \text{pre } i \neq i_0 \\ 1 & \text{pre } i, j = i_0 \\ 0 & \text{pre } i = i_0,\ i \neq j \end{cases}$$

a ohodnotenie maticou

$$\bar{z}_{ij} = \begin{cases} z_{ij} & \text{pre } i \neq i_0 \\ 0 & \text{pre } i = i_0. \end{cases}$$

Vieme, že v takto prenastavenom systéme je očakávaný výnos po n prechodoch za predpokladu, že vychádzame zo stavu i daný vzťahom

$$V_i^{(i_0)}(n) = \sum_{k=0}^{n-1} \sum_{j\in S} \bar{p}_{ij}^{(k)} \bar{c}_j. \tag{1.15}$$

Podľa prenastavení, ktoré sme vykonali, môžeme vzťah chápať ako očakávaný výnos po n prechodoch, pričom po vstupe do stavu i_0 sa kumulovanie výnosu zastaví. Očakávaný výnos do prvého vstupu do stavu i_0 ak je východiskový stav i je potom

$$V_i^{(i_0)} = \lim_{n\to\infty} V_i^{(i_0)}(n). \tag{1.16}$$

Výnos $V_{i_0}^{(i_0)}$ je zrejme 0. Pre výpočet výnosu z iného počiatočného stavu postupujme nasledovne. Označme $\mathbf{V}^{(i_0)} = \{V_i^{(i_0)}, i \in S\}$, $\mathbf{V}_{-i_0}^{(i_0)} = \{V_i^{(i_0)}, i \in S,\ i \neq i_0\}$, $\bar{\mathbf{c}}_{-i_0} = \mathbf{c}_{-i_0} = \{c_i, i \in S,\ i \neq i_0\}$ a $\bar{\mathbf{P}}_{-i_0}$ maticu, ktorá vznikne z $\bar{\mathbf{P}}$ vyškrtnutím i_0-tého stĺpca a riadku. Vieme, že platí

$$\mathbf{V}^{(i_0)} = \sum_{k=0}^{\infty} \bar{\mathbf{P}}^k \bar{\mathbf{c}}$$

Bez ujmy na všeobecnosti nech $i_0 = N$. Potom platí

$$\bar{\mathbf{P}}^k = \left[\begin{array}{ccc|c} & \bar{\mathbf{P}}^k_{-i_0} & & \begin{matrix}?\\?\\?\end{matrix} \\ \hline 0 & 0 \quad \cdots \quad 0 & & 1 \end{array}\right],$$

a $\bar{c}_N = 0$. Ďalej použijeme istú špecifickú vlastnosť matice $\bar{\mathbf{P}}^k_{-i_0}$.

Definícia 1.9 Nech je daná štvorcová matica $\mathbf{A}$ typu $n \times n$. Spektrom matice $\mathbf{A}$ nazývame množinu $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ všetkých vlastých čísel tejto matice. Spektrálnym polomerom matice $\mathbf{A}$ potom rozumieme číslo

$$\rho(\mathbf{A}) = \max_{\lambda\in\sigma(\mathbf{A})} |\lambda|$$

.

Lemma 1.10 *Nech* $\mathbf{A}$ *je všeobecne komplexná matica typu* $n \times n$*. Potom*

$$\lim_{n\to\infty} \mathbf{A}^n = 0 \Leftrightarrow \rho(\mathbf{A}) < 1$$

Matica $\bar{\mathbf{P}}_{-i_0}$ má spektrálny polomer menší než 1, a tak kombináciu lemmy 1.7 a lemmy 1.10 dostávame

$$\mathbf{V}_{-i_0}^{(i_0)} = \sum_{k=0}^{\infty} [\bar{\mathbf{P}}_{-i_0}]^k \bar{\mathbf{c}}_{-i_0} = [\mathbf{I} - \bar{\mathbf{P}}_{-i_0}]^{-1} \bar{\mathbf{c}}_{-i_0} = \bar{\mathbf{c}}_{-i_0} + \bar{\mathbf{P}}_{-i_0} \mathbf{V}_{-i_0}^{(i_0)} \tag{1.17}$$

Poznámka 1.11 *Všimnime si podobnosť medzi vzorcami* (1.12) *pre vektor očakávaných diskontovaných výnosov a* (1.17) *pre vektor očakávaných výnosov do prvého*

dosiahnutia daného stavu. Vychádzajme z matice $\mathbf{P}$ *figurujúcej vo vzorci* (1.12)*, rozšírme stavový priestor o stav* $N+1$ *a položme*

$$\bar{p}_{ij} = \begin{cases} \beta p_{ij} & \textit{pre } i,j \neq N+1 \\ (1-\beta) & \textit{pre } i \neq N+1,\ j = N+1 \\ 1 & \textit{pre } i,j = N+1 \\ 0 & \textit{pre } i = N+1,\ j \neq N+1, \end{cases}$$

$$\bar{z}_{ij} = \begin{cases} \beta^{-1} z_{ij} & \textit{pre } i,j \neq N+1 \\ 0 & \textit{pre } i \neq N+1,\ j = N+1 \\ 0 & \textit{pre } i = N+1. \end{cases}$$

Táto matica má potom stavbu matice $\bar{\mathbf{P}}$*, ktorú môžeme použiť k odvodeniu* (1.12) *pre* $i_0 = N+1$*. Je teda* $\bar{\mathbf{P}}_{-(N+1)} = \beta\mathbf{P}$ *a* $\bar{\mathbf{c}}_{-(N+1)} = \mathbf{c}$ *a tak prvky vektoru očakávaných výnosov je možné interpretovať ako očakávaný výnos do prvého dosiahnutia novo zavedeného stavu* $N+1$*.*

1.3 Spojité reťazce a semi-markovské procesy

Popis dynamického systému Markovovým reťazcom s diskrétnym časom nemusí byť pre určité situácie dostačujúci. Uvedomme si, že čas medzi jednotlivými prechodmi je v diskrétnom prípade stále rovnaký, t.j. jednotkový. Do úvahy však musíme zobrať aj úlohy, kde je nutné tento čas modelovať ako náhodný. Vhodnými nástrojmi na takéto modelovanie sú spojité Markovove reťazce, prípadne takzvané semi-Markovské procesy, ktoré sú vlastne zovšeobecnením spojitých Markovských procesov, pretože čas zotrvania v stave môže mať akékoľvek rozdelenie.

V nasledujúcom odstavci zavedieme spôsob, akým je možné takéto reťazce ohodnocovať. Nech je daný homogénny Markovov proces $\{X_t, t \geq 0\}$ so spojitým časom a diskrétnou množinou stavov S. Pre tento reťazec uvažujeme štandartne definované matice pravdepodobností prechodu $\mathbf{P}(t)$ po uplynutí $t \geq 0$ časových jednotiek a maticu intenzít prechodu $\mathbf{Q} = (q_{ij})_{i,j \in S}$. Ďalej zaveďme nasledujucí systém ohodnotenia. Prechod medzi stavom i a j realizuje zisk resp. náklad z_{ij}. Taktiež za každú časovú jednotku, po ktorú reťazec zotrvá v stave $i \in S$ obdržíme resp. zaplatíme čiastku z_i. Označme $R_i(T)$ celkový akumulovaný náhodný výnos do času T, respektíve $R_i^\beta(T), \beta > 0$ diskontovaný náhodný výnos do času T. Naším ďalším cieľom je odvodiť očakávaný výnos do času T, ak je východiskovým stavom stav $i \in S$, respektíve diskontovaný očakávaný výnos do času T. Definujeme ich ako $V_i(T) = \mathbb{E}[R_i(T)]$ a $V_i^\beta(T) = \mathbb{E}[R_i^\beta(T)]$ a zaveďme vektorové značenie $\mathbf{V}(T) = \{V_i(T),\ i \in S\}$ a $\mathbf{V}^\beta(T) = \{V_i^\beta(T),\ i \in S\}$. Pripomeňme nasledujúce základné tvrdenie.

Veta 1.12 *Nech i nieje absorbčný stav reťazca, t.j. $q_i \neq 0$. Potom má doba, po ktorú reťatec zotrvá v stave i, exponenciálne rozdelenie so strednou hodnotou $\frac{1}{q_i}$. Ďalej je pravdepodobnosť, že reťazec prejde zo stavu i najskôr do stavu j rovná $\frac{q_{ij}}{q_i}$ pre všetky $i \neq j$*

Dôkaz: v [3] veta 3.5 a veta 3.6

V prípade diskrétneho reťazca, c_i udáva očakávaný výnos za jedno časové obdobie, t.j. jeden prechod pri výstupe zo stavu i. Z predošlej vety plynie, že výraz $\frac{z_i}{q_i} + \sum_{j \in S, j \neq i} \frac{q_{ij}}{q_i} z_{ij}$ udáva očakavaný výnos do prvého výstupu (vrátane ceny za prechod) z počiatočného stavu i. Ďalej teda pre $\forall i \in S$ označíme

$$\begin{aligned} \hat{c}_i &= \frac{z_i}{q_i} + \sum_{j \in S, j \neq i} \frac{q_{ij}}{q_i} z_{ij} \\ c_i &= z_i + \sum_{j \in S, j \neq i} q_{ij} z_{ij} \end{aligned} \tag{1.18}$$

Hodnota c_i teda predstavuje sadzbu za časovú jednotku, po ktorú reťazec zotrvá v stave i. Stačí si uvedomiť, že ide vlastne o hodnotu $\hat{c}_i$ vydelenú očakávanou dobou zotrvania v stave i. Pre odvodenie vzorca na výpočet $V_i(T)$ môžeme postupovať nasledovne. Náhodný výnos $R_i(T)$ sa zrejme skladá z 2 aditívnych zložiek. Prvá je výnos, ktorý sa generuje za zotrvanie v konkrétnom stave každú časovú jednotku a

druhá je výnos za prechody. Výpočet strednej hodnoty prvej zložky je jednoduchý. Za predpokladu, že vychádzame zo stavu i definujme pre $\forall j \in S$ náhodné veličiny

$$I_j(t) = \begin{cases} 1 & \text{ak } X_t = j \\ 0 & \text{ak } X_t \neq j \end{cases}.$$

Očakávaný celkový čas, po ktorý je reťazec vychádzajúci zo stavu i v stave j až do časového bodu T je potom zrejme rovný $\mathbb{E}[\int_0^T I_j(t)dt]$. Ďalšou úpravou dostaneme

$$\mathbb{E}\left[\int_0^T I_j(t)dt\right] = \int_0^T \mathbb{E}[I_j(t)]dt = \int_0^T p_{ij}(t)dt.$$

Celkovo je teda prvá zložka rovná $\sum_{j\in S} z_j \left(\int_0^T p_{ij}(t)dt\right) = \int_0^T \left(\sum_{j\in S} p_{ij}(t)z_j\right) dt$.

K odvodeniu druhej zložky je možné využiť teóriu Poissonovych procesov a teóriu procesov obnovy. Vyslovme nasledujúce tvrdenia, ktoré sú obe dokázané v knihe [7].

Veta 1.13 *Nech $\{N_t, t \geq 0\}$ je Poissonov proces s intenzitov λ. Predpokladajme, že každý príchod je klasifikovaný ako príchod typu 1 alebo príchod typu 2, pričom príchod typu 1 nastane s pravdepodobnosťou p_1 a príchod typu 2 s pravdepodobnosťou p_2 nezávisle na všetkých ostatných príchodoch. Ďalej nech $N_1(t)$ značí počet príchodov typu 1 do času t a $N_2(t)$ počet príchodov typu 2 do času t. Potom $\{N_1(t),\ t \geq 0\}$ a $\{N_2(t),\ t \geq 0\}$ sú nezávislé Poissonove procesy postupne s intezitami λp_1 a λp_2.*

Dôkaz: v [7] veta 1.1.3

Veta 1.14 *Nech je proces obnovy $\{N_t, t \geq 0\}$, ktorý popisuje príchod do systému Poissonov s intezitou λ. Potom pre $\forall T > 0$ je*

$$\mathbb{E}[\text{ počet príchodov v intervale } (0,T),\ \text{pričom systém nájdeme v množine } B] =$$

$$= \lambda \mathbb{E}\left[\int_0^T I_B(t)dt\right]$$

Dôkaz: v [7] veta 2.4.13

Ako uvádza Tijms pre výpočet druhej zložky si uvedomme, že kedykoľvek je reťazec v stave j, tak sa výstup z tohto stavu správa ako Poissonov proces s intezitou q_j. Ďalej sa podľa vety 1.13 výstup zo stavu j do akéhokoľvek stavu $k \neq j$ správa ako Poissonov proces s intezitou $q_j \frac{q_{jk}}{q_j} = q_{jk}$. Nakoniec použijeme vetu 1.14, podľa ktorej bude počet prechodov zo stavu j do stavu k v intervale $(0,T)$ rovný $q_{jk}\int_0^T p_{ij}(t)dt$. Celkovo je teda výnos do času T plynúci z prechodov medzi stavmi rovný

$$\sum_{j\in S}\sum_{k\neq j} z_{jk}q_{jk}\int_0^T p_{ij}(t)dt = \int_0^T \left(\sum_{j\in S} p_{ij}(t)\sum_{k\neq j} z_{jk}q_{jk}\right) dt$$

Nakoniec teda máme, že pre $\forall i \in S$ je

$$V_i(T) = \int_0^T \sum_{j \in S} p_{ij}(t) c_j \, dt, \tag{1.19}$$

čo vektorovo zapíšeme ako $\mathbf{V}(T) = \int_0^T \mathbf{P}(t)\mathbf{c}dt$. Všimnime si zjavnú analógiu vzorca (1.19) s diskrétnym stavom, t.j. so vzorcom (1.5). Označme $r_i(t)$ náhodný výnos zo systému v čase t za predpokladu, že v počiatku bol reťazec v stave i. Tento výnos sa skladá z výnosu za zotrvanie reťazca v nejakom stave a prípadne aj zložky za prechod medzi stavmi. Potom pre $\forall T \geq 0$ platí $V_i(T) = \mathbb{E}[R_i(T)] = \mathbb{E}\left[\int_0^T r_i(t)\, dt\right] = \int_0^T \mathbb{E}[r_i(t)]\, dt = \int_0^T \sum_{j \in S} p_{ij}(t) c_j \, dt$, takže pre $\forall t \geq 0$ je $\mathbb{E}[r_i(t)] = \sum_{j \in S} p_{ij}(t) c_j$. Ďalej teda zrejme platí

$$V_i^\beta(T) = \mathbb{E}\left[\int_0^T e^{-\beta t} r_i(t)\, dt\right] = \int_0^T e^{-\beta t} \mathbb{E}[r_i(t)]\, dt = \int_0^T \sum_{j \in S} e^{-\beta t} p_{ij}(t) c_j \, dt \tag{1.20}$$

Definícia 1.15 Pre spojitý Markovov reťazec s ohodnotením definujeme za predpokladu $X_0 = i,\ i \in S$ celkový očakávaný diskontovaný výnos predpisom

$$V_i^\beta = \lim_{T \to \infty} V_i^\beta(T) \tag{1.21}$$

a priemerný očakávaný výnos za časovú jednotku predpisom

$$g_i = \lim_{T \to \infty} \frac{1}{T} V_i(T). \tag{1.22}$$

Zaveďme ešte vektorové označenie $\mathbf{V}^\beta = \{V_i^\beta,\ i \in S\}$.

Aj v prípade spojitých reťazcov ide o konečné ohraničené hodnoty. V prípade diskontovaného výnosu máme pre akýkoľvek východiskový stav i odhad

$$\min_{j \in S} c_j \int_0^T e^{-\beta t}(t)\, dt \leq V_i^\beta(T) = \int_0^T \sum_{j \in S} e^{-\beta t} p_{ij}(t) c_j \, dt \leq \max_{j \in S} c_j \int_0^T e^{-\beta t}(t)\, dt,$$

čo po zintegrovaní a použití limitného prechodu $T \to \infty$ dáva

$$\beta^{-1} \min_{j \in S} c_j \leq V_i^\beta \leq \beta^{-1} \max_{j \in S} c_j. \tag{1.23}$$

Pre priemerný výnos ľahko odvodíme s použitím l'Hospitaloveho pravidla a derivácie integrálu podľa hornej medze (pravdepodobnosti prechodu sú za predpokladu $\lim_{t \to 0+} p_{ij}(t) = \delta_{ij}$ rovnomerne spojité pre $\forall t \geq 0$) vzorec

$$\begin{aligned} g_i &= \lim_{T \to \infty} \frac{\int_0^T \sum_{j \in S} p_{ij}(t) c_j \, dt}{T} \\ &= \lim_{T \to \infty} \sum_{j \in S} p_{ij}(T) c_j, \end{aligned}$$

z ktorého plynie odhad

$$\min_{j\in S} c_j \leq g_i \leq \max_{j\in S} c_j.$$

Opäť odvoďme explicitné vzťahy pre výpočet. Zaveďme vektorový zápis a postupne s použitím integrácie per-partes odvoďme

$$\begin{aligned}
\mathbf{V}^\beta(T) &= \int_0^T e^{-\beta t}\mathbf{P}(t)\mathbf{c}dt \\
&= \left([-\beta^{-1}e^{-\beta t}\mathbf{P}(t)]_0^T + \int_0^T \beta^{-1}e^{-\beta t}\mathbf{P}(t)\mathbf{Q}dt\right)\mathbf{c} \\
&= \left(+\beta^{-1}\mathbf{I} - \beta^{-1}e^{-\beta T}\mathbf{P}(T)\right)\mathbf{c} + \beta^{-1}\mathbf{Q}\mathbf{V}^\beta(T)
\end{aligned}$$

Keďže je

$$\lim_{T\to\infty} -\beta^{-1}e^{-\beta T}\mathbf{P}(T) = 0$$

po dosadení nakoniec dostávame

$$\mathbf{V}^\beta = \lim_{T\to\infty} \mathbf{V}^\beta(T) = \beta^{-1}[\mathbf{c} + \mathbf{Q}\mathbf{V}^\beta] \tag{1.24}$$

Definujme maticu $\hat{\mathbf{P}} = (\hat{p}_{ij})_{i,j\in S}$ predpisom

$$\begin{aligned}
\hat{p}_{ij} &= \begin{cases} \frac{q_{ij}}{q_i} & \text{ak } q_i > 0 \\ 0 & \text{ak } q_i = 0 \end{cases}, i \neq j \\
\hat{p}_{ii} &= \begin{cases} 0 & \text{ak } q_i > 0 \\ 1 & \text{ak } q_i = 0 \end{cases}
\end{aligned}$$

Ak označíme $T_1, T_2, \ldots$ okamžiky, v ktorých dochádza k prechodu medzi stavmi reťazca a definujeme proces $\{\hat{X}_n, n \in N_0\}$ predpisom $\hat{X}_0 = X_0$ a $\hat{X}_n = X_{T_n}$, tak je tento takzvaný vnorený reťazec homogénny Markovov proces s diskrétnym časom, s maticou pravdepodobností prechodov $\hat{\mathbf{P}}$ a s množinou stavov S. Ak nebude povedané, inak budeme ďalej predpokladať, že vnorený reťazec spojitého procesu má len jedinú triedu trvalých stavov a neobsahuje absorbčný stav. K odvodeniu vzťahu pre výpočet priemerného očakávaného výnosu potom použijeme nasledujúce základné tvrdenie.

Veta 1.16 *Nech je daný spojitý Markovov reťazec $\{X_t,\ t \geq 0\}$ s konečnou množinou stavov S, ktorého vnorený reťazec má len jednu triedu trvalých stavov. Potom má reťazec jednoznačné stacionárne rozdelenie $\{\Pi_j,\ j \in S\}$, t.j. pre $\forall i, j \in S$ platí $\lim_{t\to\infty} p_{ij}(t) = \Pi_j$. Navyše platí*

$$\Pi_j = \frac{\hat{\Pi}_j/q_j}{\sum_{i\in S}\hat{\Pi}_i/q_i},$$

kde $\{\hat{\Pi}_j,\ j \in S\}$ *je stacionárne rozdelenie vnoreného reťazca.*

Dôkaz: v [7] veta 4.3.1 alebo v [3] veta 3.15 (len časť tvrdenia)

Ak má zadaný reťazec jedinú triedu trvalých stavov môžeme ľahko odvodiť hodnotu priemerného očakávaného výnosu, ktorá navyše nezávisí na východiskovom stave, nasledovne

$$g_i = \lim_{T\to\infty} \sum_{j\in S} p_{ij}(T)c_j = \sum_{j\in S} \Pi_j c_j. \tag{1.25}$$

Zamenili sme limitu a konečnú sumu a použili vetu 1.10. Za predpokladu, že má reťazec jedinú triedu trvalých stavov, budeme pre $\forall i \in S$ značiť $g = g_i$. Ak bude mať reťazec viacero tried trvalých stavov, bude priemerý výnos závisieť na tom, do ktorej triedy bude reťazec absorbovaný.

Špeciálnym druhom zovšeobecnenia spojitých Markovských procesov sú semi-Markovské procesy. Výstavbu semi-Markovského procesu zahájime nasledovne. Nech je daný akýkoľvek hodogénny Markovov reťatec s diskrétnym časom $\{\hat{X}_n, n \in N_0\}$, konečnou množinou stavov S maticou pravdepodobností prechodov $\hat{\mathbf{P}}$. Tento reťazec riadi prechody medzi jednotlivými stavmi, pričom čas zotrvania v stave i sa riadi akýmkoľvek rozdelením so strednou hodnotou τ_i. Ak pre $\forall i \in S$ za toto rozdelenie zvolíme exponenciálne so strednou hodnotou $1/q_i$ a zakážeme prechody do rovnakého stavu, t.j. $\hat{p}_{ii} = 0,\ i \in S$ dostaneme klasický spojitý Markovov proces s maticou intenzit $\mathbf{Q}$ s prvkami $q_{ij} = \bar{p}_{ij}q_i,\ i \neq j$ a $q_{ii} = -q_i$. Ak položíme čas zotrvania deterministicky rovný 1 pre všetky stavy, dostaneme klasický diskrétny Markovov reťazec. Na takomto semi-Markovskom procese zaveďme rovnaký spôsob ohodnotenia ako v prípade klasických spojitých reťazcov. Opäť bude $\hat{c}_i$ značiť očakávaný výnos do prechodu so stavu i. Z vyššie uvedených predpokladov plynie, že pre $\forall i \in S$ je

$$\hat{c}_i = \tau_i z_i + \sum_{j\in S} \hat{p}_{ij} z_{ij} \tag{1.26}$$

Veta 1.17 *Nech je daný semi-Markovský proces v zmysle predošlej definície, ktorého vnorený reťazec* $\{\hat{X}_n, n \in N_0\}$ *má len jedinú triedu trvalých stavov. Potom* $g_i = g,\ \forall i \in S$*, t.j. očakávaný priemerný výnos nezávisí na východiskovom stave. Navyše platí*

$$g = \frac{\sum_{j\in S} \hat{c}_j \hat{\Pi}_j}{\sum_{i\in S} \tau_i \hat{\Pi}_i,} \tag{1.27}$$

kde $\{\hat{\Pi}_j,\ j \in S\}$ *je stacionárne rozdelenie vnoreného reťazca.*

Dôkaz: Zovšeobecnenie vety 1.16 a výpočtu (1.25)

Ak predpokladáme, že má reťazec $\{\hat{X}_n, n \in N_0\}$ len jedinú triedu trvalých stavov, je dlhodobý priemerný výnos ohodnoteného semi-Markovského procesu rovný

$$g = \frac{K_l}{T_l},$$

kde l je pevne zvolený trvalý stav reťazca a

$$T_i \quad = \quad \text{očakávaný čas prvého vstupu do trvalého stavu } l, \text{ ak reťazec začína v stave } i$$

$$K_i \quad = \quad \text{očakávaný kumulovaný výnos do prvého vstupu do trvalého stavu } l, \text{ ak reťazec začína v stave } i$$

Tento výsledok plynie z teórie procesov obnovy s ohodnotením. Stav l je trvalý, takže v konečnom čase dôjde k vstupu reťazca do tohto stavu. Správanie reťazca je potom regeneratívne v tom zmysle, že sa proces s jednotkovou pravdepodobnosťou vráti do trvalého stavu l v konečnom čase, pričom ďalší vývoj systému má rovnaké rozdelenie ako vo východiskovom čase so stavu l a z Markovskej vlastnosti nezávisí na predošlom vývoji v systéme. V takýchto regeneratívnych procesoch je potom možné ukázať, že priemerný očakávaný výnos za časovú jednotku je rovný očakávanému výnosu za jeden cyklus delený očakávanou dĺžkov cyklu. Záujemcov odkazujeme na kapitolu 2.2 v knihe [7]. Pre očakávaný čas a výnos do vstupu do stavu l platia nasledujúce pomocné vzťahy, ktoré využijeme v ďalšom texte.

Lemma 1.18 *Nech má vnorený Markovov reťazec $\{\hat{X}_n, n \in N_0\}$ semi-Markovského procesu len jedinú triedu trvalých stavov. Pre T_i a K_i potom platí*

$$T_i \quad = \quad \tau_i + \sum_{j \neq l} p_{ij} T_j, \; i \in S \tag{1.28}$$

$$K_i \quad = \quad \hat{c}_i + \sum_{j \neq l} p_{ij} K_j, \; i \in S. \tag{1.29}$$

Dôkaz: Označme

$$\Theta_{il} \quad = \quad \text{čas prvého vstupu do trvalého stavu } l, \text{ ak reťazec začína v stave } i$$

$$\Psi_{il} \quad = \quad \text{kumulovaný výnos do prvého vstupu do trvalého stavu } l, \text{ ak reťazec začína v stave } i.$$

Do prvého vstupu do stavu l ubehne náhodný počet prechodov H. Čas a kumulovaný výnos do prvého vstupu potom môžeme rozpísať na

$$\Theta_{il} = \theta_{il}^1 + \theta_{il}^2 + \ldots \theta_{il}^H$$

$$\Psi_{il} = \psi_{il}^1 + \psi_{il}^2 + \ldots \psi_{il}^H,$$

kde θ_{il}^h značí čas ktorý uplynie od $h-1$-ho do h-teho prechodu a ψ_{il}^h značí výnos, ktorý sa nakumuluje od $h-1$-ho do h-teho prechodu. Podmienením na stav reťazca po prvom prechode postupne odvodíme

$$\begin{aligned} T_i &= \mathbb{E}[\Theta_{il}] = \mathbb{E}[\mathbb{E}[\Theta_{il}|\hat{X}_1]] = \sum_{j\in S} P(\hat{X}_1 = j)\mathbb{E}[\Theta_{il}|\hat{X}_1 = j] = \\ &= \hat{p}_{il}\tau_i + \sum_{j\neq l} \hat{p}_{ij}\mathbb{E}[\Theta_{il}|\hat{X}_1 = j] \\ &= \hat{p}_{il}\tau_i + \sum_{j\neq l} \hat{p}_{ij}(\mathbb{E}[\theta_{il}^1|\hat{X}_1 = j] + \mathbb{E}[\theta_{il}^2 + \theta_{il}^3 + \dots \theta_{il}^H|\hat{X}_1 = j]) = \\ &= \hat{p}_{il}\tau_i + \sum_{j\neq l} \hat{p}_{ij}\tau_i + \sum_{j\neq l} \hat{p}_{ij}\mathbb{E}[\Theta_{jl}] = \tau_i + \sum_{j\neq l} \hat{p}_{ij}T_j \end{aligned}$$

$$\begin{aligned} K_i &= \mathbb{E}[\Psi_{il}] = \mathbb{E}[\mathbb{E}[\Psi_{il}|\hat{X}_1]] = \sum_{j\in S} P(\hat{X}_1 = j)\mathbb{E}[\Psi_{il}|\hat{X}_1 = j] = \\ &= \hat{p}_{il}(z_i\tau_i + z_{il}) + \sum_{j\neq l} \hat{p}_{ij}\mathbb{E}[\Psi_{il}|\hat{X}_1 = j] = \\ &= \hat{p}_{il}(z_i\tau_i + z_{il}) + \sum_{j\neq l} = \hat{p}_{ij}(\mathbb{E}[\psi_{il}^1|\hat{X}_1 = j] + \mathbb{E}[\psi_{il}^2 + \psi_{il}^3 + \dots \psi_{il}^H|\hat{X}_1 = j]) = \\ &= \hat{p}_{il}(z_i\tau_i + z_{il}) + \sum_{j\neq l} \hat{p}_{ij}(z_i\tau_i + z_{ij}) + \sum_{j\neq l} \hat{p}_{ij}\mathbb{E}[\Psi_{jl}] = \hat{c}_i + \sum_{j\neq l} \hat{p}_{ij}K_j \end{aligned}$$

□

V prípade klasického spojitého reťazca majú vzorce (1.28) a (1.29) tvar

$$T_i = \frac{1}{q_i} + \sum_{j\neq l,i} \frac{q_{ij}}{q_i} T_j,\ i \in S \tag{1.30}$$

$$K_i = c_i\frac{1}{q_i} + \sum_{j\neq l,i} \frac{q_{ij}}{q_i} K_j,\ i \in S, \tag{1.31}$$

a v prípade diskrétneho Markovského procesu zase tvar

$$T_i = 1 + \sum_{j\neq l} p_{ij}T_j,\ i \in S \tag{1.32}$$

$$K_i = c_i + \sum_{j\neq l} p_{ij}K_j,\ i \in S. \tag{1.33}$$

1.4 Ďalšie vlastnosti diskrétnych reťazcov

Úpravou vzťahov (1.5) a (1.7) dostaneme pre ľubovolne zvolené reálne vektory $\mathbf{v}^\beta = \{v_i^\beta,\ i \in S\},\ \mathbf{w} = \{w_i,\ i \in S\}$ a ľubovolné reálne číslo $\bar{g}$ nasledujúce vzťahy

$$\mathbf{V}^\beta(n) = \sum_{k=0}^{n-1} \beta^k \mathbf{P}^k \mathbf{c} = \mathbf{v}^\beta - \beta^n \mathbf{P}^n \mathbf{v}^\beta + \sum_{k=0}^{n-1} \beta^k \mathbf{P}^k \boldsymbol{\varphi}^\beta, \tag{1.34}$$

$$\mathbf{V}(n) = \sum_{k=0}^{n-1} \mathbf{P}^k \mathbf{c} = n\bar{g}\mathbf{e} + \mathbf{w} - \mathbf{P}^n \mathbf{w} + \sum_{k=0}^{n-1} \mathbf{P}^k \boldsymbol{\gamma}, \tag{1.35}$$

kde sme položili

$$\boldsymbol{\varphi}^\beta = \mathbf{c} + \beta \mathbf{P} \mathbf{v}^\beta - \mathbf{v}^\beta, \tag{1.36}$$

$$\boldsymbol{\gamma} = \mathbf{c} + \mathbf{P}\mathbf{w} - \mathbf{w} - \bar{g}\mathbf{e}. \tag{1.37}$$

Pre celkový očakávaný diskontovaný výnos a očakávaný priemermy výnos potom limitným prechodom dostaneme vzťahy

$$\mathbf{V}^\beta = \lim_{n\to\infty} \mathbf{V}^\beta(n) = \mathbf{v}^\beta + \sum_{k=0}^{\infty} \beta^k \mathbf{P}^k \boldsymbol{\varphi}^\beta = \mathbf{v}^\beta + (\mathbf{I} - \beta\mathbf{P})^{-1} \boldsymbol{\varphi}^\beta, \tag{1.38}$$

$$g\mathbf{e} = \lim_{n\to\infty} \frac{\mathbf{V}(n)}{n} = \bar{g}\mathbf{e} + \frac{1}{n} \sum_{k=0}^{\infty} \mathbf{P}^k \boldsymbol{\gamma} = \bar{g}\mathbf{e} + \boldsymbol{\Pi}\boldsymbol{\gamma}. \tag{1.39}$$

Zo vzorca (1.36) okamžite plynie ekvivalencia

$$\boldsymbol{\varphi}^\beta = \mathbf{0} \Leftrightarrow \mathbf{v}^\beta = (\mathbf{I} - \beta\mathbf{P})^{-1}\mathbf{c}. \tag{1.40}$$

Okrem toho platí aj ďalšia ekvivalencia v tvare

$$\boldsymbol{\gamma} = \mathbf{0} \Leftrightarrow \bar{g}\mathbf{e} = g\mathbf{e} = \boldsymbol{\Pi}\mathbf{c}, \quad \mathbf{w} = (\mathbf{I} - \mathbf{P} + \boldsymbol{\Pi})^{-1}\mathbf{c} + k\mathbf{e}, \tag{1.41}$$

kde k je ľubovolná reálna konštanta. Platnosť ekvivalencie (1.41) overíme s využitím takzvanej fundamentálnej matice $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \boldsymbol{\Pi})^{-1}$. Ako prvé overíme, že fundamentálna matica vždy existuje a platí pre ňu vzťah $\mathbf{Z}\boldsymbol{\Pi} = \boldsymbol{\Pi}$. Stacionárna matica zrejme splňuje vťahy $\boldsymbol{\Pi}^n = \boldsymbol{\Pi},\ n \in N$ a $\mathbf{P}\boldsymbol{\Pi} = \boldsymbol{\Pi}\mathbf{P} = \boldsymbol{\Pi}$ a tak s využitím binomickej vety dostaneme pre $n \in N$ vzťah

$$(\mathbf{P} - \boldsymbol{\Pi})^n = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-1} \mathbf{P}^i \boldsymbol{\Pi}^{n-1} = \mathbf{P}^n + \sum_{i=0}^{n-1} \binom{n}{i} (-1)^{n-1} \boldsymbol{\Pi} = \mathbf{P}^n - \boldsymbol{\Pi}.$$

Je teda $\lim_{n\to\infty}(\mathbf{P}-\mathbf{\Pi})^n = \mathbf{P}^n - \mathbf{\Pi} = 0$, a tak podľa lemmy 1.7 fundamentálna matica existuje a platí pre ňu

$$Z = (\mathbf{I} - (\mathbf{P} - \mathbf{\Pi}))^{-1} = \sum_{k=0}^{\infty}(\mathbf{P} - \mathbf{\Pi})^k = \mathbf{I} + \sum_{k=1}^{\infty}(\mathbf{P}^k - \mathbf{\Pi}) \tag{1.42}$$

Vynásobením predošlého vzťahu zprava maticou $\mathbf{\Pi}$ dostaneme požadovaný vzťah $\mathbf{Z\Pi} = \mathbf{\Pi}$.

Matica $\mathbf{\Pi}$ má rovnaké riadky, a tak k vektoru $\mathbf{w}$ zrejme vždy existuje práve jeden vektor $\breve{\mathbf{w}}$, taký že $\mathbf{w} = \breve{\mathbf{w}} + \breve{k}\mathbf{e}$ a $\mathbf{\Pi}\breve{\mathbf{w}} = 0$. Kedže $\mathbf{P}$ je stochastická matica, nezáleží na tom či v (1.37) použijeme pôvodný vektor $\mathbf{w}$ alebo vektor $\breve{\mathbf{w}}$ alebo akýkoľvek iný vektor, ktorý sa líší od $\mathbf{w}$ len pripočítaním rovnakej konštanty ku každej zložke. Postupne odvodíme s využitím vzťahu (1.39) ekvivalenciu

$$\begin{aligned}
\boldsymbol{\gamma} = \mathbf{0} \quad & \Leftrightarrow \mathbf{c} + \mathbf{P}\mathbf{w} - \mathbf{w} - \bar{g}\mathbf{e} = \mathbf{0} \Leftrightarrow \mathbf{c} + \mathbf{P}\breve{\mathbf{w}} - \breve{\mathbf{w}} - \bar{g}\mathbf{e} = \mathbf{0} \Leftrightarrow \\
& \Leftrightarrow \mathbf{c} + \mathbf{P}\breve{\mathbf{w}} - \breve{\mathbf{w}} - \mathbf{\Pi}\breve{\mathbf{w}} - \bar{g}\mathbf{e} = \mathbf{0} \Leftrightarrow \mathbf{c} - \bar{g}\mathbf{e} = (\mathbf{I} - \mathbf{P} + \mathbf{\Pi})\breve{\mathbf{w}} \Leftrightarrow \\
& \Leftrightarrow \mathbf{Z}\mathbf{c} - \mathbf{Z}\bar{g}\mathbf{e} = \breve{\mathbf{w}} \Leftrightarrow \mathbf{Z}\mathbf{c} - \mathbf{Z}g\mathbf{e} = \breve{\mathbf{w}}, \bar{g}\mathbf{e} = g\mathbf{e} = \mathbf{\Pi}\mathbf{c} \\
& \Leftrightarrow \mathbf{Z}\mathbf{c} - g\mathbf{e} = \breve{\mathbf{w}}, \bar{g}\mathbf{e} = g\mathbf{e} = \mathbf{\Pi}\mathbf{c}
\end{aligned}$$

Z predošlej ekvivalencie potom plynie platnosť ekvivalencie (1.41)

Kapitola 2

Riadenie Markovových reťazcov s diskrétnym časom

2.1 Úvod

Markovove reťazce s ocenením sú vhodným nástrojom na modelovanie vývoja dynamických systémov s určitým nastavením. Veľké množstvo praktických problémov však ukázalo, že takáto štrukúra nemusí byť dostačujúca. Dynamické vývojové systémy sú totiž často podrobené kontrolám, ktorých výsledkom môže byť zmena, ktorá ovplyvní ďalšie správanie systému. Úlohou je potom nájsť vhodné akcie, t.j. určité riadenie systému tak, aby kontrolór maximalizoval (resp. minimalizoval) požadovanú účelovú funkciu, ktorou je najčastejšie priemerný očakávaný výnos alebo celkový očakávaný diskontovaný výnos.

Riadený Markovov reťazec môžeme popísať nasledujúcimi bodmi.

- Majme daný dynamický systém, ktorý je sledovaný v pravidelných časových intervaloch. V každom časovom bode $n = 0, 1, \ldots$ je systém klasifikovaný určitým stavom.
- Množinu všetkých možných stavov značme S a predpokladajme o nej, že je konečná. Bez ujmy na všeobecnosti môžeme stavy očíslovať tak, že dostaneme $S = \{1, 2, \ldots, N\}$.
- V každom časovom okamžiku $n \in N_0$ a stave $i \in S$ zvolíme rozhodnutie R_i^n, ktoré ovplyvní ďalší vývoj systému. R_i^n volíme z konečnej množiny rozhodnutí $K(i)$ viazanej k stavu i. Pre stav i uvažujme $m_i \in \mathbb{N}$ možných rozhodnutí.
- Množinu $K(i)$ uvažujeme v čase nemennú, takže v každom okamžiku máme pre stav i k dispozícii rovnaké rozhodnutia.
- Nech sú naše rozhodnutia akékoľvek, má systém Markovskú vlastnosť. Rozhodnutie R_i^n ovplyní systém v tom zmysle, že zmení pravdepodobnosti prechodov.

Prechod zo stavu i v čase n do stavu j v čase $n+1$ sa uskutoční s pravdepodobnosťou $p_{ij}(R_i^n)$, pričom zrejme musí platiť $\sum_{j\in S} p_{ij}(R_i^n) = 1$.

- S prechodom zo stavu i v čase n do stavu j v čase $n+1$ spojme zisk (náklad) $z_{ij}(R_i^n)$, ktorý sa realizuje okamžite pri výstupe zo stavu i. Zisk teda môže byť taktiež závislí na našom rozhodnutí.

Pre čas $n \in N_0$ označme $\mathbf{R}^n = (R_1^n, R_2^n, \ldots, R_N^n)'$. Našou úlohou je nájsť riadenie $\mathcal{R} = \{\mathbf{R^0}, \mathbf{R^1}, \mathbf{R^2}, \ldots\}$ t.j. vhodnú postupnosť rozhodnutí, tak aby sme maximalizovali (minimalizovali) nejakú kriteriálnu funkciu. Plánovací horizont pritom môže byť ako nekonečný tak aj konečný. Ak sa riadime podľa $\mathcal{R}$, tak pre pravdepodobnosti prechodu platí Markovská vlastnosť v tvare

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= P(X_{n+1} = j | X_n = i) \\ &= p_{ij}(R_i^n) \\ \forall n \in N_0,\ \forall\, i, j, i_{n-1}, \ldots, i_0 \in S \ \textit{pre ktoré je} & \\ P(X_{n+1} = j | X_n = i,\ X_{n-1} = i_{n-1}, \ldots,\ X_0 = i_0) &> 0 \end{aligned}$$

Všeobecne môžeme povoliť v rôznych časoch pre konkrétny stav z S rôzne rozhodnutia. V takomto prípade bude výsledný Markovov reťazec nehomogénny. Špeciálnym typom riadenia je takzvané stacionárne riadenie, ktoré priradí každému stavu $i \in S$ pevné rozhodnutie $r_i \in K(i)$. Kedykoľvek je reťazec v stave i zvolíme rozhodnutie r_i. Stacionárne riadenie reťazca zaručí jeho homogenitu. Budeme uvažovať nasledujúce všeobecné značenie. Očakávaný výnos za jedno obdobie (prechod) ak vychádzame zo stavu i v čase n, pričom zvolíme rozhodnutie R_i^n budeme v zmysle (1.1) zapisovať ako

$$c_i(R_i^n) = \sum_{j\in S} p_{ij}(R_i^n) z_{ij}(R_i^n). \tag{2.1}$$

Pre všetky stavy máme vektorový zápis

$$\mathbf{c}(R_i^n) = (c_1(R_i^n), \ldots, c_N(R_i^n))'.$$

Kedže je reťazec všeobecne nehomogénny, budeme pre akékoľvek riadenie $\mathcal{R}$ uvažovať pre $n < m$ a akýkoľvek východiskový stav i (stav v čase n) očakávaný výnos za obdobie od času n po čas m pri použití riadenie $\mathcal{R}$, ktorý označíme ako $V_i(n, m, \mathcal{R})$. Výnos po 0 krokoch je samozrejme nulový, takže $\forall n \in N_0$ definujme $V_i(n, n, \mathcal{R}) = 0$. Všeobecný vzorček pre výpočet tohto výnosu odvodíme v ďalšej kapitole.

Ak pracujeme s nekonečným plánovacím horizontom, je vo väčšine prípadov postačujúce obmedziť sa len na triedu stacionárnych riadení. Kedže je reťazec pre stacionárne riadenie homogénny, zavedieme nasledujúce zjednodušené značenie. Pre $\forall n \in N_0$ a $\forall i \in S$ môžeme značiť $R_i = R_i^n$ a $\mathbf{R} = \mathbf{R}^n = (R_1, R_2, \ldots, R_N)$. Stacionárne riadenie $\mathcal{R}$ je teda jednoznačne určené vektorom rozhodnutí $\mathbf{R}$, a tak budeme písať $\mathcal{R} = \{R_1, R_2, \ldots, R_N\}$. Pravdepodobnosti prechodu a zisk závisia na rozhodnutí

vo východiskovom stave a preto ich budeme značiť ako $p_{ij}(R_i),\ z_{ij}(R_i)$. Zaveďme tiež maticové označenie $\mathbf{P}(\mathcal{R}) = (p_{ij}(R_i))_{i,j\in S}$. V zmysle výsledkov z kapitoly 1.2 zaveďme nasledujúce označenie. Očakávaný výnos po jednom alebo viacerých obdobiach, pričom sa riadime stacionárnym riadením $\mathcal{R}$ budeme viď vzorec (1.1) a (1.6) značiť ako

$$c_i(R_i) = \sum_{j\in S} p_{ij}(R_i) z_{ij}(R_i), \tag{2.2}$$

$$V_i(n,\mathcal{R}) = c_i(R_i) + \sum_{j\in S} p_{ij}(R_i) V_j(n-1,\mathcal{R}). \tag{2.3}$$

Ak zahrnieme diskontný faktor, tak podľa (1.2) máme

$$V_i^{\beta}(n,\mathcal{R}) = c_i(R_i) + \beta \sum_{j\in S} p_{ij}(\mathcal{R}) V_j^{\beta}(n-1,\mathcal{R}). \tag{2.4}$$

Pre toto riadenie $\mathcal{R}$ ďalej podľa (1.7) respekíve (1.5) píšeme

$$V_i(n,\mathcal{R}) = \sum_{k=0}^{n-1} \sum_{j\in S} p_{ij}^{(k)}(\mathcal{R}) c_j(R_j), \tag{2.5}$$

respektíve

$$V_i^{\beta}(n,\mathcal{R}) = \sum_{k=0}^{n-1} \beta^k \sum_{j\in S} p_{ij}^{(k)}(\mathcal{R}) c_j(R_j), \tag{2.6}$$

ak uvažujeme diskontovanie. Nakoniec ak má reťazec za riadenia $\mathcal{R}$ len jedinú triedu trvalých stavov, máme pre priemerný očakávaný výnos pri použití riadenia $\mathcal{R}$ v zmysle definície 1.6 a vzorca (1.14) vyjadrenie

$$g(\mathcal{R}) = \sum_{j\in S} \Pi_j(\mathcal{R}) c_j(R_j), \tag{2.7}$$

kde $\{\Pi_i(\mathcal{R}),\ i \in S\}$ je stacionárne rozdelenie v prípade použitia riadenia $\mathcal{R}$. Pre toto staconárne rozdelenie podľa (1.13) píšeme

$$\Pi_j(\mathcal{R}) = \sum_{i\in S} \Pi_i(\mathcal{R}) p_{ij}(R_i),\ j \in S \tag{2.8}$$

$$\sum_{j\in S} \Pi_j(\mathcal{R}) = 1.$$

Pre tieto vzťahy budeme tam, kde to bude vhodné využívať vektorové zápisy v zmysle akom sme ich navrhli v kapitole 1.2. Takže napríklad píšeme $\mathbf{c}(\mathcal{R}) = \{c_i(R_i),\ i \in S\}$, $\mathbf{V}^{\beta}(n,\mathcal{R}) = \{V_i^{\beta}(n,\mathcal{R}),\ i \in S\}$, atď.

2.2 Konečný plánovací horizont

V konkrétnych úlohách nás môže zaujímať ako riadiť systém tak, aby v určitom konečnom čase dal čo najvyšší výnos resp. najnižší náklad. V prípade takéhoto konečného plánovacieho horizontu berieme do úvahy akékoľvek riadenie $\mathcal{R} = \{\mathbf{R}^0\, \mathbf{R}^1, \ldots \mathbf{R}^\nu\}$, kde ν udáva dĺžku horizontu.

Veta 2.1 *Pre očakávaný výnos $V_i(n, m, \mathcal{R}),\ n < m \leq \nu$ platí vzťah*

$$V_i(n, m, \mathcal{R}) = c_i(R_i^n) + \sum_{j \in S} p_{ij}(R_i^n) V_j(n{+}1, m, \mathcal{R})$$

Poznámka 2.2 *Zo vzorca vyplýva, že $V_i(n, n+1, \mathcal{R}) = c_i(R_i^n)$ čo je v súlade s vyššie uvedeným.*

Dôkaz: Ide len o zovšeobecnenie vety 1.1. Tentokrát sú pravdepodobnosti prechodu ako aj ocenenia určené riadením $\mathcal{R}$, takže pri tomto pevnom riadeni môže byť reťazec nehomogénny. Vďaka Markovskej vlastnosti môžeme bez ujmy na všeobecnosti položiť východiskový čas $n = 0$. Zvoľme ľubovolné $i \in S,\ m \in N_0$ a riadenie $\mathcal{R}$. Uskutočníme m prechodov pričom realizácia javu

$$[X_0 = i, X_1 = i_1, \ldots, X_m = i_m]$$

dá výnos

$$z_{ii_1}(R_i^0) + z_{i_1 i_2}(R_{i_1}^1) + \cdots + z_{i_{m-1} i_m}(R_{i_{m-1}}^{m-1}).$$

Postupným podmieňovaním s využitím Markovskej vlastnosti odvodíme podobne ako (1.3), že táto realizácia nastave s pravdepodobnosťou

$$p_i p_{ii_1}(R_i^0)\, p_{i_1 i_2}(R_{i_1}^1) \ldots p_{i_{m-1} i_m}(R_{i_{m-1}}^{m-1}),$$

kde $p_i = P(X_0 = i)$. Očakávaný výnos je potom podľa výpočtu (1.4), s prihliadnutím na nehomogenitu možné zapísať a ako

$$V_i(0, m, \mathcal{R}) = c_i(R_i^0) + \sum_{i_1 \in S} p_{ii_1}(R_i^0) V_{i_1}(1, m, \mathcal{R}).$$

□

Veta 2.3 *Pre $\forall i \in S,\ n = 0, \ldots, m-1$ definujme rekurzívne*

$$\hat{V}_i(n, m) = \max_{r_i \in K(i)} \left[c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) \hat{V}_j(n{+}1, m) \right]$$

kde $\hat{V}_i(m, m) = 0$. Potom pre $\forall \mathcal{R},\ \forall i \in S,\ n = 0, \ldots, m-1$ platí

$$V_i(n, m, \mathcal{R}) \leq \hat{V}_i(n, m)$$

Poznámka 2.4 *Definovali sme* $\hat{V}_i(m-1,m) = \max_{r_i \in K(i)} c_i(r_i)$. *Ide vlastne o maximálny stredný výnos za jedno obdobie pričom vychádzame zo stavu* i.

Dôkaz: Zvoľme ľubovolné $i \in S$ a prípustné riadenie $\mathcal{R}$. Postupujeme rekurzívne. Pre $n = m-1$ máme:

$$V_i(m-1,m,\mathcal{R}) = c_i(R_i^{m-1}) \leq \max_{r \in K(i)} c_i(r) = \hat{V}_i(m-1,m)$$

Nech tvrdenie vety platí pre nejaké $n = 1, 2, \ldots, m-1$. Ukážeme, že potom platí aj pre $n-1$

$$\begin{aligned} V_i(n-1,m,\mathcal{R}) &= c_i(R_i^{n-1}) + \sum_{j \in S} p_{ij}(R_i^{n-1}) V_j(n,m,\mathcal{R}) \\ &\leq c_i(R_i^{n-1}) + \sum_{j \in S} p_{ij}(R^{n-1}) \hat{V}_j(n,m) \\ &\leq \max_{r_i \in K(i)} \left[c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) \hat{V}_j(n,m) \right] = \hat{V}_i(n-1,m) \end{aligned}$$

□

Predchádzajúca veta nám vlastne dáva návod ako rekurzívne od konca horizontu napočítať optimálne riadenie maximalizujúce výnosy $V_i(n,m,\mathcal{R})$. Tento takzvaný Bellmanov princíp optimality pre úlohu stochastického dynamického programovania ilustrujme nasledujúcim jednoduchým príkladom.

Problém výrobcu cukroviniek: Výrobca cukroviniek plánuje pravidelne sledovať predajnosť svojho výrobku, ktorý sa bude predávať ešte 6 období (6 kvartálov). Výrobok je potom podľa predajnosti v každom sledovanom období hodnotený ako

- úspešný - stav 1
- normálny - stav 2
- neúspešný - stav 3

Predpokladajme, že predajnosť v danom období závisí len na predajnosti v minulom období a že celá dynamika môže byť popísaná Markovovým reťazcom. V každom čase a stave má výrobca možnosť zvoliť jednu z nasledujúcich akcií:

- Rozhodnutie 1 - nevykonať žiadnu akciu
- Rozhodnutie 2 - investovať do marketingu
- Rozhodnutie 3 - zvýšiť ceny

(a) Pravdepodobnosti prechodov

$p_{ij}(1)$		j		
		1	2	3
i	1	0,35	0,55	0,10
	2	0,15	0,60	0,25
	3	0,10	0,40	0,50

(b) Ohodnotenie prechodov

$z_{ij}(1)$		j		
		1	2	3
i	1	9	7	2
	2	6	2	-1
	3	-1	-4	-12

Tabulka 2.1: Rozhodnutie 1 - nevykonať akciu

(a) Pravdepodobnosti prechodov

$p_{ij}(2)$		j		
		1	2	3
i	1	0,45	0,50	0,05
	2	0,30	0,50	0,20
	3	0,15	0,60	0,25

(b) Ohodnotenie prechodov

$z_{ij}(2)$		j		
		1	2	3
i	1	6	5	-1
	2	4	-0,25	-4
	3	-2	-6	-14

Tabulka 2.2: Rozhodnutie 2 - investovať do marketingu

(a) Pravdepodobnosti prechodov

$p_{ij}(3)$		j		
		1	2	3
i	1	0,30	0,50	0,20
	2	0,10	0,50	0,40
	3	0,00	0,35	0,65

(b) Ohodnotenie prechodov

$z_{ij}(3)$		j		
		1	2	3
i	1	12,15	9	2
	2	7	3	-1
	3	0	-4	-13

Tabulka 2.3: Rozhodnutie 3 - zvýšiť ceny

V tabuľkách 2.1,2.2 a 2.3 sú pravdepodobnosti prechodov a ocenenie prechodov pri voľbe rôznych akcií. Možnosť investovať do marketingu, predstavuje zvýšenie pravdepodobnosti, že sa výrobok bude predávať lepšie, avšak za cenu vyšších výrobných nákladov. Naopak zvýšenia cien spôsobí zvýšenie pravdepodobností prechodov do horších kategórií, avšak výnosy sa zvýšia. Stredné výnosy po jednom období vypočítané podľa (2.1) sú v nasledujúcej tabulke 2.4.

$c_i(r)$		r		
		1	2	3
i	1	7,2	5,15	8,55
	2	1,85	0,27	1,80
	3	-7,7	-7,40	-9,85

Tabulka 2.4: Očakávané výnosy po 1 prechode

Pozrime sa najprv na situáciu ktorá, by nastala ak by sme reťazec neriadili, t.j.

nevykonávali žiadnu akciu (rozhodnutie 1). Podľa (2.5) vypočítajme stredné výnosy po n obdobiach. Zaokrúhlené výsledky sú zapísané v nasledujúcej tabuľke 2.5.

n	1	2	3	4	5	6
$V_1(n)$	7,20	9,97	10,84	11,06	11,05	10,97
$V_2(n)$	1,85	2,11	2,09	2,01	1,90	1,80
$V_3(n)$	-7,70	-10,09	-10,90	-11,23	-11,41	-11,54

Tabulka 2.5: Očakávané výnosy

V ďalšej tabuľke 2.6 môžeme vidieť zaokrúhlené napočítané hodnoty $\hat{V}_i(m-n,m)$ a k nim príslušné rozhodnutia r_i^n pre plánovací horizont 6 období. Rozhodnutia $r_i^n,\ i=1,2,3,\ n=1,\ldots,6$ predstavujú optimálne riadenie systému.

n	1	2	3	4	5	6
$\hat{V}_1(m-n,m)$	8,55	10,55	11,54	12,42	13,31	14,19
r_1	3	3	3	1	1	1
$\hat{V}_2(m-n,m)$	1,85	2,39	3,27	4,15	5,03	5,91
r_2	1	1	2	2	2	2
$\hat{V}_3(m-n,m)$	-7,40	-6,86	-6,10	-5,23	-4,36	-3,48
r_3	2	2	2	2	2	2

Tabulka 2.6: Očakávané výnosy

Vidíme, že ak sa nachádzame v stave 1, t.j. výrobok je úspešný, je výhodné spočiatku nevykonávať žiadnu akciu a až ako sa blíži ukončenie predaja, je vhodné zvýšiť ceny. V prípade zlého predaja výrobku je vhodné vždy investovať do reklamy. Ak má výrobok neutrálne výsledky, je spočiatku vhodné investovať do reklamy. Ako sa blíži záver predaja, tak sa od tejto stratégie upustí a nevykoná sa žiadna akcia. V tomto prípade sa nám zrejme investícia do reklamy už nevráti. Každopádne pre akýkoľvek východiskový stav dostaneme lepší hospodársky výsledok ako keby sme reťazec neriadili.

2.3 Nutná a postačujúca podmienka optimality pre nekonečný časový horizont

V tomto odstavci odvodíme nutnú a postačujúcu podmienku pre optimálne riadenie, ak je kritériom diskontovaný alebo priemerný očakávaný výnos. Použitý prístup je rozšírením vzťahov odvodených v kapitole 1.4. Ako už bolo uvedené v poznámke 1.4 pri optimalizácii systémov, ktoré sú popísané Markovskými procesmi, sa stačí obmedziť na Markovské riadenia, takže rozhodnutie v určitom čase závisí iba na aktuálnom stave systému.

Pre vektory očakávaných diskontovaných výnosov po n prechodoch pri použití stacionárneho riadenie $\mathcal{R} = \{R_1, R_2, \ldots, R_N\}$ a ľubovolného riadenia $\bar{\mathcal{R}} = \{\bar{\mathbf{R}}^1, \bar{\mathbf{R}}^2, \ldots\}$ môžeme rozšírením vzťahu (1.34) na nehomogénny prípad pri voľbe $v_i^\beta = V_i^\beta(\mathcal{R}),\ i \in S$ písať

$$\begin{aligned}\mathbf{V}^\beta(\bar{\mathcal{R}}, n) &= \sum_{k=0}^{n-1} \beta^k \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\mathbf{c}(\bar{\mathbf{R}}^k) \\ &= \mathbf{V}^\beta(\mathcal{R}) - \beta^n \prod_{l=0}^{n} \mathbf{P}(\bar{\mathbf{R}}^l)\mathbf{V}^\beta(\mathcal{R}) + \sum_{k=0}^{n-1} \beta^k \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\boldsymbol{\varphi}^\beta(\bar{\mathbf{R}}^k, \mathcal{R}),\end{aligned}$$

kde

$$\boldsymbol{\varphi}^\beta(\bar{\mathbf{R}}^k, \mathcal{R}) = \mathbf{c}(\bar{\mathbf{R}}^k) + \beta\mathbf{P}(\bar{\mathbf{R}}^k)\mathbf{V}^\beta(\mathcal{R}) - \mathbf{V}^\beta(\mathcal{R}). \tag{2.9}$$

Pre nekonečný časový horizont teda limitným prechodom dostávame

$$\mathbf{V}^\beta(\bar{\mathcal{R}}) = \lim_{n\to\infty} \mathbf{V}^\beta(\bar{\mathcal{R}}, n) = \mathbf{V}^\beta(\mathcal{R}) + \sum_{k=0}^{\infty} \beta^k \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\boldsymbol{\varphi}^\beta(\bar{\mathbf{R}}^k, \mathcal{R}). \tag{2.10}$$

Ak nájdeme stacionárne riadenie $\tilde{\mathcal{R}} = \{\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N\}$ také, že pre akékoľvek prípustné stacionárne riadenie $\hat{\mathcal{R}} = \hat{\mathbf{R}} = \{\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_N\}$ je $\boldsymbol{\varphi}^\beta(\hat{\mathbf{R}}, \tilde{\mathcal{R}}) \leq 0$ (teda $\varphi_i^\beta(\hat{R}_i, \tilde{\mathcal{R}}) \leq 0,\ i \in S$, všimnime si, že $\varphi_i^\beta(\tilde{R}_i, \tilde{\mathcal{R}}) = 0,\ i \in S$), potom stacionárne riadenie $\tilde{\mathcal{R}}$ maximalizuje očakávaný diskontovaný výnos pre každý počiatočný stav. Z rovnice (2.10) taktiež plynie, že riadenie (nestacionárne) $\bar{\mathcal{R}}$ je optimálne práve vtedy keď $\boldsymbol{\varphi}^\beta(\bar{\mathbf{R}}^k, \tilde{\mathcal{R}}) = 0$ pre $\forall k \in N$. Algoritmické postupy pre nájdenie optimálneho riadenia sú popísané v kapitole 2.4.1, 2.4.2 a 2.8.

V prípade, že kritériom je priemerný výnos, dostaneme rozšírenim (1.35) na nehomogénny prípad pri voľbe $\bar{g} = g(\mathcal{R})$ a $w_i = w_i(\mathcal{R}),\ i \in S$ vzťah

$$\begin{aligned}\mathbf{V}(\bar{\mathcal{R}}, n) &= \sum_{k=0}^{n-1} \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\mathbf{c}(\bar{\mathbf{R}}^k) \\ &= ng\mathbf{e} + \mathbf{w}(\mathcal{R}) - \prod_{l=0}^{n} \mathbf{P}(\bar{\mathbf{R}}^l)\mathbf{w}(\mathcal{R}) + \sum_{k=0}^{n-1} \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\boldsymbol{\gamma}^\beta(\bar{\mathbf{R}}^k, \mathcal{R}),\end{aligned}$$

kde

$$\boldsymbol{\gamma}^{\beta}(\bar{\mathbf{R}}^k, \mathcal{R}) = \mathbf{c}(\bar{\mathbf{R}}^k) + \mathbf{P}(\bar{\mathbf{R}}^k)\mathbf{w}(\mathcal{R}) - \mathbf{w}(\mathcal{R}) - g(\mathcal{R})\mathbf{e}$$

Takže pre priemerný výnos za nekonečný časový horizont máme

$$\begin{aligned} g(\bar{\mathcal{R}})\mathbf{e} = \lim_{n\to\infty} \frac{\mathbf{V}(\bar{\mathcal{R}}, n)}{n} = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\mathbf{c}(\bar{\mathbf{R}}^k) = \\ = g(\mathcal{R})\mathbf{e} + \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\boldsymbol{\gamma}^{\beta}(\bar{\mathbf{R}}^k, \mathcal{R}). \end{aligned} \tag{2.11}$$

Ak nájdeme stacionárne riadenie $\tilde{\mathcal{R}} = \{\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N\}$, také že pre akékoľvek prípustné stacionárne riadenie $\hat{\mathcal{R}} = \hat{\mathbf{R}} = \{\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_N\}$ je $\boldsymbol{\gamma}(\hat{\mathbf{R}}, \tilde{\mathcal{R}}) \leq 0$ (teda $\gamma_i(\hat{R}_i, \tilde{\mathcal{R}}) \leq 0,\ i \in S$, zrejme $\gamma_i(\tilde{R}_i, \tilde{\mathcal{R}}) = 0,\ i \in S$), potom stacionárne riadenie $\tilde{\mathcal{R}}$ maximalizuje očakávaný priemerny výnos. Z rovnice (2.11) taktiež plynie, že riadenie (nestacionárne) $\bar{\mathcal{R}}$ je optimálne práve vtedy keď

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \prod_{l=0}^{k-1} \mathbf{P}(\bar{\mathbf{R}}^l)\boldsymbol{\gamma}^{\beta}(\bar{\mathbf{R}}^k, \mathcal{R}) = \mathbf{0}.$$

Algoritmické postupy pre nájdenie optimálneho riadenia sú popísané v kapitole 2.6.1, 2.6.2.

2.4 Optimálny diskontovaný výnos

2.4.1 Algoritmus policy iteration

Pri hľadaní riadenia s optimálnou hodnotou očakávaného diskontovaného výnosu musíme samozrejme počítať s nekonečným plánovacím horizontom a preto sa stačí obmedziť na triedu stacionárnych riadení. Nech je teda dané ľubovoľné stacionárne riadenie $\mathcal{R}$. Očakávaný diskontovaný výnos za n období pri použití riadenia $\mathcal{R}$ budeme značiť ako $V_i^\beta(n, \mathcal{R})$ a celkový očakávaný diskontovaný výnos ako $V_i^\beta(\mathcal{R})$. Kedže používame stacionárne riadenie, je skúmaný reťazec homogénny a tak môžeme použiť výsledky z kapitoly 1.2. Pre výpočet očakávaného diskontovaného výnosu po n obdobiach pri použití riadenia $\mathcal{R}$ teda platí rekurentný vzťah (1.2), t.j.

$$V_i^\beta(n, \mathcal{R}) = c_i(R_i) + \beta \sum_{j \in S} p_{ij}(\mathcal{R}) V_j^\beta(n-1, \mathcal{R}). \tag{2.12}$$

Navyše podľa dôsledku 1.8 platí

$$\mathbf{V}^\beta(\mathcal{R}) = (\mathbf{I} - \beta \mathbf{P}(\mathcal{R}))^{-1} \mathbf{c}(\mathcal{R}) \tag{2.13}$$

Nasledujúca veta je dôležitou súčasťou algoritmického postupu na hľadanie stacionárneho riadenia, ktoré maximalizuje dlhodobý diskontovaný výnos.

Veta 2.5 *Nech pre dané čísla $v_i^\beta,\ i \in S$ platí*

$$c_i(R_i) + \beta \sum_{j \in S} p_{ij}(R_i) v_j^\beta \geq v_i^\beta,\ \forall i \in S.$$

Potom pre dlhodobý diskontovaný výnos platí

$$V_i^\beta(\mathcal{R}) \geq v_i^\beta,\ i \in S.$$

Dôkaz Pre $\forall i \in S$ platí

$$\begin{aligned} V_i^\beta(\mathcal{R}) - v_i^\beta &\geq c_i(R_i) + \beta \sum_{j \in S} p_{ij}(R_i) V_j^\beta(\mathcal{R}) - \left(c_i(R_i) + \beta \sum_{j \in S} p_{ij}(R_i) v_j^\beta \right) \\ &\geq \beta \sum_{j \in S} p_{ij}(R_i)(V_j^\beta(\mathcal{R}) - v_j^\beta). \end{aligned}$$

Čo môžeme maticovo zapísať ako

$$\mathbf{V}^\beta(\mathcal{R}) - \mathbf{v}^\beta \geq \beta \mathbf{P}(\mathcal{R})(\mathbf{V}^\beta(\mathcal{R}) - \mathbf{v}^\beta),$$

alebo ekvivalentne ako

$$(\mathbf{I} - \beta \mathbf{P}(\mathcal{R}))(\mathbf{V}^\beta(\mathcal{R}) - \mathbf{v}^\beta) \geq \mathbf{0}. \tag{2.14}$$

K dokončeniu dôkazu stačí obe strany nerovnosti zľava vynásobiť maticou

$$(\mathbf{I} - \beta\mathbf{P}(\mathcal{R}))^{-1} = \sum_{k=0}^{\infty}[\beta\mathbf{P}(\mathcal{R})]^k > \mathbf{0}$$

□

Predchádzajúce výsledky nám dávajú možnosť zostaviť nasledujúci algoritmus.

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $V_i^{\beta}(\mathcal{R}),\ i \in S$ sústavy lineárnych rovníc

  $$v_i^{\beta} \quad = \quad c_i(R_i) + \beta \sum_{j \in S} p_{ij}(R_i) v_j^{\beta},\ i \in S,$$

  o neznámych $v_i^{\beta},\ i \in S$. Riešenie $V_i^{\beta}(\mathcal{R}),\ i \in S$ je podľa (2.13) očakávaný diskontovaný výnos pri použití riadenia $\mathcal{R}$.

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

  $$c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i) V_j^{\beta}(\mathcal{R}),\ i \in S. \tag{2.15}$$

  Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizije výraz (2.15), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S$. Z vety 2.5 potom plynie, že diskontovaný výnos pre nové riadenie bude vyšší.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1, pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$. Algoritmus musí skonvergovať, pretože stacionárnych riadení je konečný počet a každé nové riadenie dáva vyššiu hodnotu očakávaného diskontovaného výnosu.

V predchádzajúcom algoritme teda iteráciou aktuálneho riadenia postupne dospejeme k riadeniu optimálnemu. Takýto typ algoritmu sa v literátúre najčastejšie pomenúva policy-iteration. Všimnime si, že v každej iterácii je v kroku 1 nutné vyriešiť sústavu rovníc o N neznámych, čo môže byť v situáciách veľkého stavového priestoru niekedy výpočtovo náročné. Preto sa niekedy používajú iné druhy aproximatívnych algoritmov, ktoré sú výpočtovo vhodnejšie. Takéto algoritmy často nenájdu priamo optimálne riadenie, ale len riadenie, za ktorého je kriteriálna funkcia dostatočne blízko optimálnej.

2.4.2 Algoritmus value iteration

Z predošlého policy-iteration algoritmu vyplýva, že existujú jednoznačné čísla $\tilde{V}_i^\beta, \in S$ splňujúce

$$\tilde{V}_i^\beta = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i) \tilde{V}_j^\beta \right\}, \ i \in S \tag{2.16}$$

Stačí si uvedomiť, že v poslednom a predposlednom kroku algoritmu sa vygeneruje rovnaké (optimálne) riadenie $\tilde{\mathcal{R}}$. V poslednom kroku algoritmu teda riadenie $\tilde{\mathcal{R}}$ maximalizuje výraz (2.15) a tak platí (2.16), pričom je $\tilde{V}_i^\beta = V_i^\beta(\tilde{\mathcal{R}}), \ i \in S$ v zmysle kroku 2 Policy iteration algoritmu. Ďalej je zrejmé, že každé riadenie $\mathcal{R}$, ktoré maximalizuje (2.16), je optimálne vhľadom na celkový očakávaný diskontovaný výnos, pričom maximálny diskontovaný výnos pri výstupe zo stavu i je jednoznačne daný konštantou $\tilde{V}_i^\beta$. Aproximatívny, v literatúre označovaný value iteration algoritmus, potom môžeme zostaviť na základe nasledujúceho tvrdenia.

Veta 2.6 *Pre postupne napočítané hodnoty*

$$u_i^\beta(n) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i) u_j^\beta(n-1) \right\}, \tag{2.17}$$

kde $u_i^\beta(0) = 0, \ i \in S$ *platí pre* $\forall i \in S$*, že*

$$\lim_{n \to \infty} u_i^\beta(n) = \tilde{V}_i^\beta.$$

Dôkaz: Nech $\mathcal{R}_n = \{R_1, \ldots, R_N\}$ značí stacionárne riadenie, také že pre $\forall i \in S$ práve rozhodnutie R_i maximalizuje v n-tej iterácii pravú stranu (2.17). S využitím vektorového značenia teda platí

$$\mathbf{u}(n+1) = \max_{\mathcal{R}} [\mathbf{c}(\mathcal{R}) + \beta \mathbf{P}(\mathcal{R}) \mathbf{u}(n)] = \mathbf{c}(\mathcal{R}_n) + \beta \mathbf{P}(\mathcal{R}_n) \mathbf{u}(n) \tag{2.18}$$

Zvoľme ľubovolné riadenie optimálne vzhľadom k celkovému diskontovanému výnosu a označme ho $\tilde{\mathcal{R}}$. Vzorec (2.16) potom môžeme prepísať do vektorového tvaru

$$\tilde{\mathbf{V}}^\beta = \max_{\mathcal{R}} [\mathbf{c}(\mathcal{R}) + \beta \mathbf{P}(\mathcal{R}) \tilde{\mathbf{V}}^\beta] = \mathbf{c}(\tilde{\mathcal{R}}) + \beta \mathbf{P}(\tilde{\mathcal{R}}) \tilde{\mathbf{V}}^\beta. \tag{2.19}$$

Pre $n \in N$ definujme $\mathbf{y}(n) = \mathbf{u}(n) - \tilde{\mathbf{V}}^\beta$. Podľa (2.18) a (2.19) je

$$\mathbf{y}(n+1) = \mathbf{c}(\mathcal{R}_n) + \beta \mathbf{P}(\mathcal{R}_n) \mathbf{u}(n) - [\mathbf{c}(\tilde{\mathcal{R}}) + \beta \mathbf{P}(\tilde{\mathcal{R}}) \tilde{\mathbf{V}}^\beta]$$

Ďalej zrejme platí

$$\begin{aligned} \mathbf{y}(n+1) &\leq \mathbf{c}(\mathcal{R}_n) + \beta \mathbf{P}(\mathcal{R}_n) \mathbf{u}(n) - [\mathbf{c}(\mathcal{R}_n) + \beta \mathbf{P}(\mathcal{R}_n) \tilde{\mathbf{V}}^\beta] \\ &= \beta \mathbf{P}(\mathcal{R}_n) [\mathbf{V}(n) - \tilde{\mathbf{V}}^\beta] = \beta \mathbf{P}(\mathcal{R}_n) \mathbf{y}(n) \end{aligned}$$

$$\begin{aligned}\mathbf{y}(n+1) &\geq \mathbf{c}(\tilde{\mathcal{R}}) + \beta\mathbf{P}(\tilde{\mathcal{R}})\mathbf{u}(n) - [\mathbf{c}(\tilde{\mathcal{R}}) + \beta\mathbf{P}(\tilde{\mathcal{R}})\tilde{\mathbf{V}}^{\beta}] \\ &= \beta\mathbf{P}(\tilde{\mathcal{R}})[\mathbf{u}(n) - \tilde{\mathbf{V}}^{\beta}] = \beta\mathbf{P}(\tilde{\mathcal{R}})\mathbf{y}(n).\end{aligned}$$

Postupným dosadením nakoniec dostaneme

$$\beta^n(\mathbf{P}(\tilde{\mathcal{R}}))^n\mathbf{y}(1) \leq \mathbf{y}(n+1) \leq \beta^n\mathbf{P}(\mathcal{R}_n)\dots\mathbf{P}(\mathcal{R}_1)\mathbf{y}(1)$$

Je teda $\lim_{n\to\infty}\mathbf{y}(n) = \mathbf{0}$, takže $\forall i \in S$ platí $\lim_{n\to\infty} u_i^{\beta}(n) = \tilde{V}_i^{\beta}$. □

Z predošlej vety a (2.16) v skutočnosti plynie, že stacionárne riadenie $\mathcal{R}_n$, ktorého rozhodnutia $R_i, i \in S$ maximalizujú v n-tej iterácii pravú stranu (2.17) postupne pre $n \to \infty$ konverguje k optimálnemu riadeniu. Formálnym dôkazom tohto tvrdenia je nasledujúca lemma, ktorá navyše poskytuje odpoveď na otázku ako zostaviť ukončovacie pravidlo algoritmického nápočtu, t.j. kedy je posledné vygenerované riadenie už dostatočne blízklo optimálnemu.

Lemma 2.7 *Pre $\forall n \in N$ označme $B_n = \max_{i\in S}\{|u_i^{\beta}(n) - u_i^{\beta}(n-1)|\}$. Potom pre akékoľvek $j \in S$ platí*

$$\begin{aligned}|\tilde{V}_j^{\beta} - u_j^{\beta}(n)| &\leq \frac{\beta}{1-\beta}B_n, \\ |V_j^{\beta}(\mathcal{R}_n) - u_j^{\beta}(n)| &\leq \frac{\beta}{1-\beta}B_n\end{aligned}$$

Dôkaz: Zvoľme ľubovolné $n \in N$. Z využitím vektorového zápisu z predošlej vety máme

$$\mathbf{u}^{\beta}(n) = \max_{\mathcal{R}}\{\mathbf{c}(\mathcal{R}) + \beta\mathbf{P}(\mathcal{R})\mathbf{u}^{\beta}(n-1)\} = \mathbf{c}(\mathcal{R}_n) + \beta\mathbf{P}(\mathcal{R}_n)\mathbf{u}^{\beta}(n-1). \quad (2.20)$$

Zo vzťahu (2.20) potom plynie platnosť nasledujúcich dvoch nerovností.

$$\mathbf{u}^{\beta}(n+1) - \mathbf{u}^{\beta}(n) \leq \beta\mathbf{P}(\mathcal{R}_{n+1})[\mathbf{u}^{\beta}(n) - \mathbf{u}^{\beta}(n-1)]$$

$$\mathbf{u}^{\beta}(n+1) - \mathbf{u}^{\beta}(n) \geq \beta\mathbf{P}(\mathcal{R}_n)[\mathbf{u}^{\beta}(n) - \mathbf{u}^{\beta}(n-1)],$$

ktoré po ďalšej úprave dávajú

$$\max_{i\in S}\{u_i^{\beta}(n+1) - u_i^{\beta}(n)\} \leq \beta\max_{i\in S}\{u_i^{\beta}(n) - u_i^{\beta}(n-1)\} \quad (2.21)$$

$$\min_{i\in S}\{u_i^{\beta}(n+1) - u_i^{\beta}(n)\} \geq \beta\min_{i\in S}\{u_i^{\beta}(n) - u_i^{\beta}(n-1)\}. \quad (2.22)$$

Využitím nerovnosti (2.21) si môžeme všimnúť, že platí

$$\begin{aligned} u_j^\beta(n+2) - u_j^\beta(n) &= u_j^\beta(n+2) - u_j^\beta(n+1) + u_j^\beta(n+1) - u_j^\beta(n) \\ &\leq \beta(\beta+1) \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\} \end{aligned}$$

Indukciou dokážeme platnosť vzťahu

$$u_j^\beta(n+m) - u_j^\beta(n) \leq \beta \frac{1-\beta^m}{1-\beta} \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\}, \tag{2.23}$$

pre $m \in N$. Z predošleho už vieme, že tvrdenie platí pre $m = 1, 2$. Nech tvrdenie platí pre m. Dokazujme platnosť pre $m+1$. Využijeme rozpis

$$u_j^\beta(n+m+1) - u_j^\beta(n) = [u_j^\beta(n+m+1) - u_j^\beta(n+m)] + [u_j^\beta(n+m) - u_j^\beta(n)]$$

Pre prvý člen máme podľa (2.21) vzťah

$$u_j^\beta(n+m+1) - u_j^\beta(n+m) \leq \beta^{m+1} \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\}.$$

Pre druhý člen použijeme (2.23). Celkovo teda dostaneme

$$u_j^\beta(n+m+1) - u_j^\beta(n) \leq \beta \frac{1-\beta^{m+1}}{1-\beta} \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\},$$

čo je požadovaná nerovnosť. Limitným prechodom pre $m \to \infty$ dostaneme s využitím vety 2.6

$$\tilde{V}_j^\beta - u_j^\beta(n) \leq \frac{\beta}{1-\beta} \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\}$$

Rovnakým postupom s využitím (2.22) dostaneme

$$\tilde{V}_j^\beta - u_j^\beta(n) \geq \frac{\beta}{1-\beta} \min_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\}$$

Spojením predošlých dvoch nerovností dosnaneme prvé tvrdenie lemmy.

Druhú nerovnosť dokážeme nasledovne. Zvoľme ľubovolné $n \in N$, položme $\bar{\mathbf{u}}^\beta(n) = \mathbf{u}^\beta(n)$, značme $\mathcal{R}_n = (R_{n,1}, R_{n,2}, \ldots, R_{n,N})$ a napočítavajme

$$\bar{u}_i^\beta(n+m) = c_i(R_{n,i}) + \beta \sum_{j \in S} p_{ij}(R_{n,i}) \bar{u}_j^\beta(n+m-1).$$

Hodnoty po n-tom kroku teda už nenapočítavame maximalizáciou cez všetky prípustné rozhonutia, ale priamo dosadzujme n-té riadenie. Rovnakým postupom ako v prvej časti dôkazu odvodíme nerovnosť

$$\bar{u}_j^\beta(n+m) - u_j^\beta(n) \leq \beta \frac{1-\beta^m}{1-\beta} \max_{i \in S}\{u_i^\beta(n) - u_i^\beta(n-1)\}, \tag{2.24}$$

respektíve opačnú nerovnosť, v ktorej figuruje minimalizácia rozdielu napočítaných hodnôt. Keďže je tentokrát $\lim_{m\to\infty} \bar{u}_j^{\beta}(n+m) = V_j^{\beta}(\mathcal{R}_n)$, limitný prechodom a obdobnými úpravami ako v prvej časti dôkazu dospejeme k požadovanej nerovnosti. □

Dôsledkom lemmy je platnosť nerovnosti

$$\max_{i\in S}\{|\tilde{V}_i^{\beta} - V_i^{\beta}(\mathcal{R}_n)|\} \le \max_{i\in S}\{|\tilde{V}_i^{\beta} - u_i^{\beta}(n)|\} + \max_{i\in S}\{|u_i^{\beta}(n) - V_i^{\beta}(\mathcal{R}_n)|\} \le \frac{2\beta B_n}{1-\beta}.$$

Z vety 2.6 plynie, že $B_n \to 0$ a preto $\max_{i\in S}\{|\tilde{V}_i^{\beta} - V_i^{\beta}(\mathcal{R}_n)|\} \to 0$. Navyše nám veta umožňuje zostaviť vhodné ukončovacie pravidlo.

Predošlé výsledky môžeme zhrnúť do nasledujúceho value-iteration algoritmu

- Krok 0 - Inicializácia
  Položíme $u_i^{\beta}(0) = 0,\ i \in S$. Ďalej nastavíme $n := 1$ a postúpime na samotný iteračný algoritmus.

- Krok 1 - Value iteration
  Pre každý stav $i \in S$ spočítame

  $$u_i^{\beta}(n) = \max_{r_i\in K(i)} \left\{ c_i(r_i) + \beta \sum_{j\in S} p_{ij}(r_i) u_j^{\beta}(n-1) \right\}.$$

  Nech $\mathcal{R}$ je stacionárne riadenie ktorého príslušné rozhodnutia $R_i,\ i \in S$ maximalizujú pravú stranu výrazu.

- Krok 2 - Hranice
  Spočítame hranicu

  $$B_n = \max_{i\in S}\{|u_i(n) - u_i(n-1)|\}.$$

- Krok 3 - Ukončovacie pravidlo

  Ak

  $$0 \le B_n \le \frac{(1-\beta)\epsilon}{2\beta}$$

  kde $\epsilon > 0$ je predom určený koeficient presnosti, algoritmus skončí a za optimálne riadenie vezmeme $\mathcal{R}$. V opačnom prípade položíme $n := n+1$ a pokračujeme krokom 1.

Po ukončení algoritmu teda máme

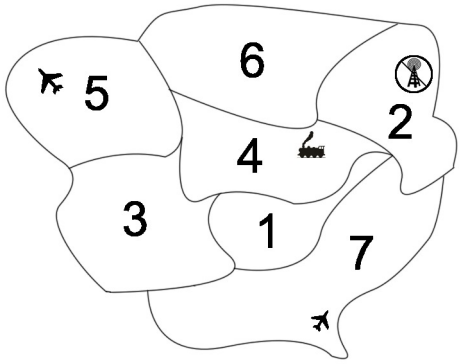
$$\max_{i\in S}\{|\tilde{V}_i^{\beta} - V_i^{\beta}(\mathcal{R})|\} \le \frac{2\beta B_n}{1-\beta} \le \epsilon, \tag{2.25}$$

2.4.3 Taxikárov problém

Nasledujúci problém je rozšírením úlohy (viac stavov a stretégií) z knihy [1]. Taxikár pracuje vo veľkomeste, ktoré sa skladá zo 7 oblastí. V každej oblasti môže zvoliť z niekoľkých stratégií poskytovania svojej služby. Celkovo má k dispozícií nasledujúce stratégie

- 1. stratégia:
  Krúžiť mestskou časťou až kým si ho neprivolá okoloidúci.
- 2. stratégia:
  Zájsť k najbližšiemu parkovaciemu miestu pre taxikárov a čakať na zákazníka v rade.
- 3. stratégia:
  Zastaviť na mieste a čakať na zavolanie dispečera cez rádio
- 4.stratégia
  Zájsť na letisko alebo železničnú stanicu a čakať na zákazníka

Mesto má 2 letiská nachádzajúce sa v mestských častiach 5 a 7 a jednu železničnú stanicu nachádzajúcu sa v mestkej časti 4. V mestkej časti 2 nie je k dispozícií dispečing. Pre danú mestskú oblasť a zvolenú stratégiu máme pre každú cielovú oblasť k dispozícií pravdepodobnosť, že zákazník bude chcieť odviesť práve do tejto cieľovej oblasti. K tejto pravdepodobnosti je priradené ocenenie, ktoré zahrňuje príjem taxikára, náklady spojené s cestou a náklady spojené so zvolenou stratégiou. Pravdepodobnosti prechodu a očakávané výnosy za jeden prechod uvedené v tabuľke 2.7 teda závisia na zvolenej stratégii, pretože voľba stratégie spôsobí, že sa taxikár zameriava na rozdielnu populáciu zákazníkov. Úlohou je nájsť riadenie, t.j. pre každú oblasť vybrať stratégiu tak, aby sme maximalizovali celkový diskontovaný výnos taxikára. Uvažujeme diskontný faktor $\beta = 0.9$.



Obrázek 2.1: Plán mestských oblastí

i	r	$p_{ij}(r)$							$c_i(r)$
		$j=$ 1	2	3	4	5	6	7	
1	1	0.2	0.15	0.15	0.3	0.05	0.1	0.05	20.15
	2	0.1	0.2	0.2	0.05	0.2	0.15	0.1	22.25
	3	0.2	0.15	0.1	0.1	0.15	0.1	0.2	21.4
2	1	0.35	0.1	0.05	0.25	0.05	0.05	0.15	12.15
	2	0.15	0.1	0.25	0.15	0.2	0.025	0.125	16.625
3	1	0.2	0.1	0.1	0.25	0.1	0.15	0.1	13.75
	2	0.1	0.15	0.30	0.15	0.1	0.1	0.1	14.55
	3	0.4	0.025	0.05	0.35	0.05	0.025	0.1	13.325
4	1	0.3	0.05	0.05	0.4	0.025	0.15	0.025	16.275
	2	0.2	0.075	0.15	0.15	0.3	0.025	0.1	14
	3	0.1	0.1	0.15	0.25	0.25	0.125	0.025	16.675
	4	0.025	0.2	0.175	0.05	0.2	0.2	0.15	14.075
5	1	0.1	0.025	0.175	0.2	0.3	0.1	0.1	11.525
	2	0.05	0.05	0.175	0.2	0.25	0.15	0.125	16.45
	3	0.1	0.1	0.1	0.25	0.15	0.05	0.25	17
	4	0.15	0.15	0.05	0.25	0.025	0.075	0.3	18.85
6	1	0.05	0.05	0.075	0.45	0.15	0.2	0.025	12.3
	2	0.15	0.1	0.1	0.2	0.2	0.1	0.15	20.25
	3	0.05	0.05	0.05	0.3	0.25	0.15	0.15	15.55
7	1	0.15	0.15	0.15	0.2	0.05	0.1	0.2	12.55
	2	0.25	0.15	0.175	0.225	0.05	0.05	0.1	11.575
	3	0.1	0.2	0.1	0.15	0.25	0.1	0.1	15.45
	4	0.025	0.175	0.05	0.2	0.4	0.1	0.05	19.775

Tabulka 2.7: Pravdepodobnosti prechodu

Ako prvý použime policy-iteration algoritmus. Výsledky pre jednotlivé iterácie môžeme vidieť v tabuľke 2.8, ktorá obsahuje riešenie sústavy rovníc z kroku 1, t.j. očakávané diskontované výnosy pri použití aktuálne iterovaného riadenia a tabuľke 2.9, ktorá obsahuje vygenerované riadenie z kroku 2. Vidíme, že algoritmus skonverguje po 4 iteráciách

Iter.	Vstupné riadenie	Riešenie sústavy rovníc						
		V_1^β	V_2^β	V_3^β	V_4^β	V_5^β	V_6^β	V_7^β
1	$\{1,1,1,1,1,1,1\}$	158.47	151.17	151.60	155.66	148.04	149.84	149.98
2	$\{1,2,3,1,4,2,4\}$	180.21	176.63	174.14	176.95	179.64	180.98	180.41
3	$\{2,2,1,3,4,2,4\}$	185.52	179.67	177.51	179.88	182.59	183.97	183.14
4	$\{2,2,3,1,4,2,4\}$	186.40	180.60	178.60	181.24	183.52	184.89	184.06

Tabulka 2.8: Riešenie - policy iteration

Vyriešme úlohu aj pomocou value-iteration algoritmu s voľbou $\epsilon = 0.5$, to znamená že po ukončení algoritmu dostaneme riadenie, ktorého očakávaný diskontovaný výnos pri výstupe z akéhokoľvek stavu sa od maximálneho líši najviac o 0.5. Iterované hodnoty a generované riadenie nájdeme v tabuľke 2.10. Všimnime si, že iterované

Iter.	Vypočítané riadenie						
stav=	1	2	3	4	5	6	7
0	1	1	1	1	1	1	1
1	1	2	3	1	4	2	4
2	2	2	1	3	4	2	4
3	2	2	3	1	4	2	4
4	2	2	3	1	4	2	4

Tabulka 2.9: Riešenie - policy iteration

hodnoty skutočne konvergujú k maximálnemu diskontovanému výnosu. Algoritmus konverguje veľmi pomaly a zastaví sa až po 63 iteráciách. Rýchlejšie konvergujúci value-iteration algoritmus si predstavíme v kapitole 2.8.

It. (n)	Vypočítané hodnoty							Riadenie	B_n
	$u_1^\beta(n)$	$u_2^\beta(n)$	$u_3^\beta(n)$	$u_4^\beta(n)$	$u_5^\beta(n)$	$u_6^\beta(n)$	$u_7^\beta(n)$		
1	22.25	16.63	14.55	16.68	18.85	20.25	19.78	$\{2,2,2,3,4,2,4\}$	22.25
2	38.52	32.72	30.70	33.29	35.64	36.95	36.05	$\{2,2,3,1,4,2,4\}$	16.785
3	53.28	47.47	45.48	48.12	50.37	51.76	50.95	$\{2,2,3,1,4,2,4\}$	14.897
10	122.72	116.92	114.93	117.57	119.84	121.21	120.39	$\{2,2,3,1,4,2,4\}$	7.075
20	164.19	158.39	156.40	159.04	161.31	162.68	161.86	$\{2,2,3,1,4,2,4\}$	2.467
40	183.70	177.90	175.90	178.54	180.82	182.19	181.36	$\{2,2,3,1,4,2,4\}$	0.860
63	186.16	180.36	178.36	181.00	183.28	184.65	183.82	$\{2,2,3,1,4,2,4\}$	0.0266

Tabulka 2.10: Riešenie - value iteration

Optimálne správanie taxikára je teda

- mestská časť 1: zájsť k parkovaciemu miestu a čakať
- mestská časť 2: zájsť k parkovaciemu miestu a čakať
- mestská časť 3: čakať na zavolanie rádiom
- mestská časť 4: krúžiť mestskou časťou
- mestská časť 5: zájsť na letisko
- mestská časť 6:zájsť k parkovaciemu miestu a čakať
- mestská časť 7: zájsť na letisko

2.5 Tranzientné programovanie

Opäť uvažujeme množinu stavov $S = \{1, 2, \ldots, N\}$. Pre ľubovolne zvolený stav $i_0 \in S$ chceme náš dynamický systém riadiť tak, aby sme maximalizovali očakávaný výnos do prvého vstupu do stavu i_0. Pre akékoľvek prípustné stacionárne riadenie $\mathcal{R}$ predpokladajme, že je matica $\mathbf{P}(\mathcal{R})$ nerozložitelná a položme

$$\bar{p}_{ij}(\mathcal{R}) = \begin{cases} p_{ij}(\mathcal{R}) & \text{pre } i \neq i_0 \\ 1 & \text{pre } i, j = i_0 \\ 0 & \text{pre } i = i_0,\ i \neq j, \end{cases}$$

$$\bar{z}_{ij}(\mathcal{R}) = \begin{cases} z_{ij}(\mathcal{R}) & \text{pre } i \neq i_0 \\ 0 & \text{pre } i = i_0. \end{cases}$$

Opäť využijeme značenie z kapitoly 1.2 pričom budeme parametrizovať nejakým stacionárnym riadením. Pre akékoľvek stacionárne riadenie $\mathcal{R}$ potom píšeme

$$V_i^{(i_0)}(n, \mathcal{R}) = \sum_{k=0}^{n-1} \sum_{j \in S} \bar{p}_{ij}^{(k)}(\mathcal{R}) \bar{c}_j(\mathcal{R}),$$

$$V_i^{(i_0)}(\mathcal{R}) = \lim_{n \to \infty} V_i^{(i_0)}(n, \mathcal{R}).$$

Navyše vieme, že pre toto riadenie je $V_{i_0}^{(i_0)}(\mathcal{R}) = 0$ (na riadení v stave i_0 teda zrejme nezáleží) a podľa (1.17) platí

$$\mathbf{V}_{-i_0}^{(i_0)}(\mathcal{R}) = [\mathbf{I} - \bar{\mathbf{P}}_{-i_0}(\mathcal{R})]^{-1} \bar{\mathbf{c}}_{-i_0}(\mathcal{R}) = \bar{\mathbf{c}}_{-i_0}(\mathcal{R}) + \bar{\mathbf{P}}_{-i_0}(\mathcal{R}) \mathbf{V}_{-i_0}^{(i_0)}(\mathcal{R})$$

Algoritmus na nájdenie optimálneho riadenia typu policy iteration môžeme zostavaviť nasledovne.

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $v_i^{(i_0)}(\mathcal{R}),\ i \in S \setminus \{i_0\}$ sústavy lineárnych rovníc:

$$v_i^{(i_0)} \quad = \quad c_i(R_i) + \sum_{j \in S \setminus \{i_0\}} \bar{p}_{ij}(R_i) v_j^{(i_0)},\ i \in S \setminus \{i_0\}.$$

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

$$c_i(r_i) + \sum_{j \in S \setminus \{i_0\}} \bar{p}_{ij}(r_i) v_j^{i_0}(\mathcal{R}),\ i \in S \setminus \{i_0\}. \tag{2.26}$$

Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizije výraz (2.26), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S \setminus \{i_0\}$. Pre úplnosť kladieme $\overline{R}_{i_0} = R_{i_0}$ (zrejme by sme sa stavom i_0 nemuseli zaoberať). Nové riadenie $\overline{\mathcal{R}}$ má potom pre akýkoľvek východiskový stav vyššiu alebo rovnakú hodnotu výnosu do dosiahnutia stavu i_0. Dôkaz tohto tvrdenia je analogický dôkazu vety 2.5.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1 pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$.

2.6 Optimálny priemerný výnos

2.6.1 Algoritmus policy iteration

V tejto kapitole odvodíme policy iteration algoritmus na hľadanie optimálneho riadenia, ktoré maximalizuje priemerný očakávaný výnos za časovú jednotku. Pracujeme s nekonečným časovým horizintom, a preto sa opäť stačí obmedziť na triedu stacionárnych riadení. Začneme tvdením, ktoré nám umožňuje posudzovať, ktoré z dvojice riadení je lepšie.

Veta 2.8 *Nech je dané stacionárne riadenie $\mathcal{R}$, pre ktoré má príslušný Markovov ret'azec len jedinú triedu trvalých stavov. Nech pre dané čísla g a $v_i,\ i \in S$ platí*

$$c_i(R_i) - g + \sum_{j \in S} p_{ij}(R_i) v_j \geq v_i,\ \forall i \in S. \tag{2.27}$$

Potom dlhodobý priemerný výnos splňuje

$$g(\mathcal{R}) \geq g. \tag{2.28}$$

Ostrá nerovnosť bude v (2.28) práve vtedy, ak za riadenia $\mathcal{R}$ existuje trvalý stav l taký, že v (2.27) bude pre $i = l$ ostrá nerovnosť.

Dôkaz: Predpokladáme, že reťazec má pri použití riadenia $\mathcal{R}$ jedinú triedu trvalých stavov, a tak máme k dispozícii jednoznačné stacionárne rozdelenie $\{\Pi_j(\mathcal{R}),\ j \in S\}$. Prenásobením rovnice (2.27) pre každé i výrazom $\Pi_i(\mathcal{R})$ a následným sčítaním cez i dostaneme

$$\sum_{i \in S} \Pi_i(\mathcal{R}) c_i(R_i) - g + \sum_{i \in S} \Pi_i(\mathcal{R}) \sum_{j \in S} p_{ij}(R_i) v_j \geq \sum_{i \in S} \Pi_i(\mathcal{R}) v_i,$$

pričom ostrú nerovnosť dostaneme práve vtedy ak máme v (2.27) ostrú nerovnosť pre nejaký stav l, taký že $\Pi_l(\mathcal{R}) > 0$. Dosadením podľa (2.7) s využitím (2.8) dostaneme

$$g(\mathcal{R}) - g + \sum_{j \in S} \Pi_j(\mathcal{R}) v_j \geq \sum_{i \in S} \Pi_i(\mathcal{R}) v_i.$$

Odčítaním výrazu $\sum_{i \in S} \Pi_i(\mathcal{R}) v_i$ z oboch strán rovnice obdržíme tvrdenie vety. □

Ak predpokladáme, že má reťazec za použitia $\mathcal{R}$ len jedinú triedu trvalých stavov, je totiž dlhodobý priemerný výnos rovný

$$g(\mathcal{R}) = \frac{K_l(\mathcal{R})}{T_l(\mathcal{R})},$$

kde l je pevne zvolený trvalý stav reťazca pri použití riadenia $\mathcal{R}$ a

$T_i(\mathcal{R})$ = očakávaný čas prvého vstupu do trvalého stavu l,
ak reťazec začína v stave i a je použité riadenie $\mathcal{R}$

$K_i(\mathcal{R})$ = očakávaný kumulovaný výnos do prvého vstupu do trvalého stavu l,
ak reťazec začína v stave i a je použité riadenie $\mathcal{R}$.

Vyššie popísané porovnajte s úvahami v poslednom odstavci kapitoly 1.3.

Definícia 2.9 Nech je dané stacionárne riadenie $\mathcal{R}$, pre ktoré má príslušný Markovov reťazec len jedinú triedu trvalých stavov. Potom definujeme relatívne hodnoty

$$w_i(\mathcal{R}) = K_i(\mathcal{R}) - g(\mathcal{R})T_i(\mathcal{R}),\ i \in S, \tag{2.29}$$

pričom zaveďme konvenciu, že za referenčný stav l zvoľme najväčší trvalý stav v zmysle očíslovania v rámci S. Pripomeňme, že pracujeme s $S = \{1, \ldots, N\}$.

Veta 2.10 *Nech je dané stacionárne riadenie $\mathcal{R}$, pre ktoré má príslušný Markovov reťazec len jedinú triedu trvalých stavov. Priemerný výnos $g(\mathcal{R})$ a relatívne hodnoty $w_j(\mathcal{R}),\ j \in S$ sú potom riešením sústavy lineárnych rovníc*

$$v_i = c_i(R_i) - g + \sum_{j \in S} p_{ij}(R_i)v_j,\ i \in S \tag{2.30}$$

o neznámych g a $v_j,\ j \in S$.

Pre túto sústavu platí, že ak sú čísla g a $v_j,\ j \in S$ jej riešením, tak pre nejakú konštantu c platí

$$g = g(\mathcal{R}),\ v_j = w_j(\mathcal{R}) + c,\ j \in S.$$

Pre ľubovolne zvolený stav s má sústava spolu s podmienkou $v_s = 0$ jednoznačné riešenie.

Poznámka 2.11 *Riešitelnosť rovnice (2.30) je možné ukázať pomocou fundamentálnej matice tak ako sme to urobili v kapitole 2.3. Algoritmus však dokážeme s využitím procesov obnovy s ohodnotením. Hlavná myšlinka nasledujúceho dôkazu je z knihy [7] a uvádzame ho tu z toho dôvodu, že v kapitole 3 ukážeme prepis tohto dôkazu pre spojité riadené procesy.*

Dôkaz: Ako prvé dokážeme, že $g(\mathcal{R})$ a $w_j(\mathcal{R}),\ j \in S$ sú riešením sústavy (2.30). Zvoľme ľubovolné $i \in S$. Dosaďme $g(\mathcal{R})$ a $w_j(\mathcal{R}),\ j \in S$ do pravej strany príslušnej rovnice. Dostaneme

$$c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(R_i)w_j(\mathcal{R})$$

Ďalej podľa definície relatívnej hodnoty a faktu, že $w_l = 0$ upravíme na

$$c_i(R_i) - g(\mathcal{R}) + \sum_{j \neq l} p_{ij}(R_i)[K_j(\mathcal{R}) - g(\mathcal{R})T_j(\mathcal{R})].$$

K dokončeniu dôkazu stačí použiť dôsledok lemmy 1.18. Podľa (1.32) a (1.33) môžeme pre stacionárne riadenie $\mathcal{R}$ písať

$$\begin{aligned} T_i(\mathcal{R}) &= 1 + \sum_{j \neq l} p_{ij}(R_i)T_j(\mathcal{R}),\ i \in S \\ K_i(\mathcal{R}) &= c_i(R_i) + \sum_{j \neq l} p_{ij}(R_i)K_j(\mathcal{R}).\ i \in S \end{aligned}$$

Pravá strana sa potom rovná

$$\begin{aligned} c_i(R_i) + \sum_{j \neq l} p_{ij}(R_i)K_j(\mathcal{R}) - g(\mathcal{R}) \left[1 + \sum_{j \neq r} p_{ij}(R_i)T_j(\mathcal{R})\right] = \\ = K_i(\mathcal{R}) - g(\mathcal{R})T_i(\mathcal{R}) = w_i(\mathcal{R}) \end{aligned}$$

Nech g a $v_j,\ j \in S$ je ľubovolné riešenie sústavy. Indukciou overíme, že pre $n \in N_0$ platí

$$v_i = V_i(n, \mathcal{R}) - ng + \sum_{j \in S} p_{ij}^{(n)}(\mathcal{R})v_j,\ i \in S. \tag{2.31}$$

Za platnosti tohto vzťahu s využitím toho, že

$$\begin{aligned} \lim_{n \to \infty} V_i(n, \mathcal{R})/n = g(\mathcal{R}), \\ \lim_{n \to \infty} \frac{1}{n} \sum_{j \in S} p_{ij}^{(n)}(\mathcal{R})v_j = 0, \end{aligned}$$

je už jednoduché ukázať, že $g = g(\mathcal{R})$. Stačí len obe strany (2.31) vydeliť n a využiť limitný prechod pre $n \to \infty$.

Dokazujme teda indukciou platnosť (2.31). Pre $n = 1$ je tvrdenie zrejmé, keďže g, $v_j,\ j \in S$ je riešením sústavy. Predpokladjme, že (2.31) platí pre n. Dokazujeme platnosť pre $n + 1$. Postupným dosadením za $V_i(n, \mathcal{R})$ podľa (2.5) a za v_j podľa (2.30) dostaneme

$$\begin{aligned} v_i &= \sum_{k=0}^{n-1} \sum_{j \in S} c_j(R_j)p_{ij}^{(k)}(\mathcal{R}) - ng + \sum_{j \in S} p_{ij}^{(n)}(\mathcal{R})v_j \\ &= \sum_{k=0}^{n-1} \sum_{j \in S} c_j(R_j)p_{ij}^{(k)}(\mathcal{R}) - ng + \sum_{j \in S} p_{ij}^{(n)}(\mathcal{R}) \left[c_j(R_j) - g + \sum_{k \in S} p_{jk}(R_j)v_k\right] \\ &= \sum_{k=0}^{n} \sum_{j \in S} c_j(R_j)p_{ij}^{(k)}(\mathcal{R}) - (n+1)g + \sum_{k \in S} \sum_{j \in S} \left[p_{ij}^{(n)}(\mathcal{R})p_{jk}(R_j)\right] v_k,\ i \in S \end{aligned}$$

V poslednom člene sme zamenili poradie konečných súm. S využitím Chapman-Kolmogorovej rovnosti dostaneme dokazované tvrdenie.

Ďalej dokazujeme, že pre nejakú konštantu c platí $v_i = w_i(\mathcal{R}) + c,\ i \in S$. Nech g' a $v'_i,\ i \in S$ je nejaké iné riešenie sústavy. Podľa predchádzajúcej časti dôkazu platí $g = g' = g(\mathcal{R})$, a tak z (2.30) dostaneme

$$v_i - v'_i = \sum_{j \in S} p_{ij}^{(n)}(\mathcal{R})[v_j - v'_j],\ i \in S,\ n \in N_0.$$

Sčítaním rovnice pre $n = 1, \ldots, m$ a následným vydelením m dostaneme

$$\frac{1}{m}\sum_{n=1}^{m}[v_i - v'_i] = \frac{1}{m}\sum_{n=1}^{m}\sum_{j \in S} p_{ij}^{(n)}(\mathcal{R})[v_j - v'_j],\ i \in S,\ m \in N_0.$$

Upravíme a zameníme poradie sumácie. Stále sa jedná o konečné sumy.

$$v_i - v'_i = \sum_{j \in S}\left(\frac{1}{m}\sum_{n=1}^{m} p_{ij}^{(n)}(\mathcal{R})\right)[v_j - v'_j],\ i \in S,\ m \in N_0$$

Limitným prechodom pre $m \to \infty$ obdržíme

$$v_i - v'_i = \sum_{j \in S} \Pi_j(\mathcal{R})[v_j - v'_j],\ i \in S.$$

Pravá strana tejto rovnice už nezávisí na i. Tým je tvrdenie dokázané.

Pre posledné tvrdenie vety stačí overiť, že pre ľubovolnú konštantu c je $v_i = w_i(\mathcal{R}) + c,\ i \in S$ riešením sústavy. Pre každé $i \in S$ s využitím faktu, že $\sum_{j \in S} p_{ij}(R_i) = 1$ zrejme platí

$$\begin{aligned} w_i(\mathcal{R}) + c &= c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(R_i) w_j(\mathcal{R}) + c \\ &= c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(R_i)(w_j(\mathcal{R}) + c). \end{aligned}$$

Stačí len odčítať konštantu c z oboch strán čím dostaneme platnú rovnosť. Podmienka $v_s = 0$ potom jednoznačne určí konštantu $c = -w_s(\mathcal{R})$. Ukázali sme teda, že v tomto prípade existuje jednoznačné riešenie. □

Na základe predchádzajúcich tvrdení zostavíme algoritmus policy iteration.

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $g(\mathcal{R}),\ v_i(\mathcal{R})$ sústavy lineárnych rovníc:

$$\begin{aligned} v_i &= c_i(R_i) - g + \sum_{j \in S} p_{ij}(R_i) v_j,\ i \in S, \\ v_s &= 0, \end{aligned}$$

  kde s je ľubovolne zvolený stav. Korektnosť kroku zaručuje veta 2.10.

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

$$c_i(r_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R}). \tag{2.32}$$

  Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizije výraz (2.32), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S$. Stačí si uvedomiť, že maximum výrazu (2.32) je pre $\forall i \in S$ väššie alebo rovné $v_i(\mathcal{R})$ a použiť vetu 2.8, podľa ktorej je $g(\mathcal{R}) \leq g(\overline{\mathcal{R}})$.

- Informácia o maximálnom priemernom výnose
  Spočítame

$$L_n = \min_{i \in S} \max_{r_i \in K(i)} \{c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R}) - v_i(\mathcal{R})\}$$

$$U_n = \max_{i \in S} \max_{r_i \in K(i)} \{c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R}) - v_i(\mathcal{R}),\}$$

  kde n značí poradie aktuálnej iterácie. Ak maximálny priemerný výnos označíme ako $\tilde{g}$ tak viď poznámka 2.17 platí

$$L_n \leq \tilde{g} \leq U_n,$$

  pričom tieto medze tvoria monotónnu postupnosť. Využijeme to hlavne pri veľkých úlohách, pre ktoré môže byť algoritmus časovo náročný.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1 pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$

Poznámka 2.12 *Namiesto vypočítaných hodnôt* $c_i(r_i) = \sum_{j \in S} p_{ij}(r_i) z_{ij}(r_i),\ i \in S,\ r_i \in K(i)$ *ktoré dávajú očakávaný výnos za jeden prechod, môžeme v algoritme pracovať priamo s danými hodnotami* $c_i(r_i)$*, ktoré potom interpretujeme ako výnos resp. náklad spojený s voľbou rozhodnutia* r_i *v stave* i*.*

Konvergencia Policy iteration algoritmu nieje až taká zrejmá ako sa na prvý pohľad môže zdať. Uvedomme si, že môže nastať situácia, že pre dve rozdielne prípustne stacionárne riadenia máme rovnakú hodnotu dlhodobého priemerného výnosu. Nemôže sa teda algoritmus zacykliť medzi takýmito dvoma riadeniami? Negatívnu odpoveď na túto otázku zaručuje nasledujúce tvrdenie.

Veta 2.13 *Nech* $\mathcal{R}$ *a* $\overline{\mathcal{R}}$ *sú stacionárne riadenia vygenerované algoritmom, ktoré sú bezprostredne za sebou, také že* $\mathcal{R} \neq \overline{\mathcal{R}}$*. Potom platí*
i) $g(\mathcal{R}) < g(\overline{\mathcal{R}})$ *alebo*
ii) $g(\mathcal{R}) = g(\overline{\mathcal{R}})$ *a* $w_i(\mathcal{R}) \leq w_i(\overline{\mathcal{R}})$ *pre* $\forall i \in S$*, pričom je nerovnosť ostrá aspoň pre jeden stav* i*.*

Poznámka 2.14 *Stacionárnych riadení je konečný počet a podľa vety sa v každom kroku buď zvýši dlhodobý priemerný výnos alebo aspoň jedna relatívna hodnota, čo zaručuje požadovanú konvergenciu.*

Dôkaz: Podľa kroku 2 v Policy iteration algoritme je

$$c_i(\overline{R}_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(\overline{R}_i) w_j(\mathcal{R}) \geq w_i(\mathcal{R}),\ i \in S \tag{2.33}$$

a tak je podľa vety 2.8 $g(\mathcal{R}) \leq g(\overline{\mathcal{R}})$. Predpokladajme teda, že $g(\mathcal{R}) = g(\overline{\mathcal{R}})$. Podľa druhej časti vety 2.8 (aplikovanej na riadenie $\overline{\mathcal{R}}$, kde $g = g(\mathcal{R})$ a $v_i = v_i(\mathcal{R}),\ i \in S$) a konvencie, ktorú sme zaviedli v kroku 2 dostávame implikáciu

$$g(\mathcal{R}) = g(\overline{\mathcal{R}}) \Rightarrow R_i = \overline{R}_i,\ i \in I(\overline{\mathcal{R}})$$

Trvalé stavy tvoria nerozložitelnú množinu a tak predchádzajúce implikuje $I(\mathcal{R}) = I(\overline{\mathcal{R}})$, t.j. za oboch riadení máme rovnakú množinu trvalých stavov. Podľa definície relatívnych hodnôt 2.9 a konvencie, ktorú sme v nej zaviedli nahliadneme, že platí $w_i(\mathcal{R}) = w_i(\overline{\mathcal{R}}),\ i \in I(\overline{\mathcal{R}})$. Ak vychádzame z trvalého stavu, hodnoty relatívnych hodnôt už totiž nezávisia na riadení reťazca v prechodných stavoch a za referenčný stav volíme v oboch prípadoch rovnaký stav. Uvedomme si, že všetky stavy nemôžu byť za predpokladu $g(\mathcal{R}) = g(\overline{\mathcal{R}})$ trvalé, pretože by sme dostali spor s $\mathcal{R} \neq \overline{\mathcal{R}}$. Predtým ako sa pozrieme na prechodné stavy uvážme nasledujúce. Podľa vety 2.10 platí

$$w_i(\mathcal{R}) = c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} p_{ij}(R_i) w_j(\mathcal{R}),\ i \in S. \tag{2.34}$$

Rovnice (2.34) zrejme paltia aj pre $\overline{\mathcal{R}}$. Odčítaním i-tej rovnice (2.34) pre riadenie $\mathcal{R}$ od i-tej rovnice (2.34) pre riadenie $\overline{\mathcal{R}}$ dostaneme pre $\forall i \in S$

$$w_i(\overline{\mathcal{R}}) - w_i(\mathcal{R}) = c_i(\overline{R}_i) + \sum_{j \in S} p_{ij}(\overline{R}_i) w_j(\overline{\mathcal{R}}) - c_i(R_i) - \sum_{j \in S} p_{ij}(R_i) w_j(\mathcal{R}). \tag{2.35}$$

Definujme $\Gamma_i(\mathcal{R},\overline{\mathcal{R}}) = c_i(\overline{R}_i) - q_i(R_i) + \sum_{j\in S}[p_{ij}(\overline{R}_i) - p_{ij}(R_i)]w_j(\mathcal{R}), i \in S$.
Po dosadení (2.34) za $w_i(\mathcal{R})$ v (2.33) l'ahko nahliadneme, že $\Gamma_i(\mathcal{R},\overline{\mathcal{R}}) \geq 0$ pre $\forall i \in S$.
Rovnice (2.35) prepíšme pomocou $\Gamma_i(\mathcal{R},\overline{\mathcal{R}})$ na tvar

$$w_i(\overline{\mathcal{R}}) - w_i(\mathcal{R}) = \Gamma_i(\mathcal{R},\overline{\mathcal{R}}) + \sum_{j\in S} p_{ij}(\overline{R}_i)[w_j(\overline{\mathcal{R}}) - w_j(\mathcal{R})].$$

Zaoberajme sa teda len prechodnými stavmi. Definujme vektory

$$\mathbf{u} = \{w_i(\overline{\mathcal{R}}) - w_i(\mathcal{R}),\ i \notin I(\overline{\mathcal{R}})\} \text{ a } \mathbf{\Gamma}(\mathcal{R},\overline{\mathcal{R}}) = \{\Gamma_i(\mathcal{R},\overline{\mathcal{R}}),\ i \notin I(\overline{\mathcal{R}})\}$$

Pre prechodné stavy môžeme rovnice (2.35) prepísať na vektorový tvar

$$\mathbf{u} = \mathbf{\Gamma}(\mathcal{R},\overline{\mathcal{R}}) + \widetilde{P}(\overline{\mathcal{R}})\mathbf{u},$$

kde $\widetilde{P}(\overline{\mathcal{R}})$ je matica so spektrálnym polomerom menším ako 1, ktorá vznikne z $P(\overline{\mathcal{R}})$ vyškrtnutím trvalých stavov. Môžeme teda písať $\mathbf{u} = [\mathbf{I} - \widetilde{P}(\overline{\mathcal{R}})]^{-1}\mathbf{\Gamma}(\mathcal{R},\overline{\mathcal{R}})$, čo je ekvivalentné s $\mathbf{u} = \sum_{n=0}^{\infty}\widetilde{P}^n(\overline{\mathcal{R}})\mathbf{\Gamma}(\mathcal{R},\overline{\mathcal{R}})$. Pretože matice $\widetilde{P}^n(\overline{\mathcal{R}}),\ n \in N$ aj vektor $\mathbf{\Gamma}(\mathcal{R},\overline{\mathcal{R}})$ obsahujú len nezáporné prvky, musí nutne platiť $\mathbf{u} > \mathbf{0}$. Rovnosť nastať nemôže, pretože by sme dostali spor s predpokladom $\mathcal{R} \neq \overline{\mathcal{R}}$. Záverom teda dostávame, že za predpokladu $g(\mathcal{R}) = g(\overline{\mathcal{R}})$ je $w_i(\overline{\mathcal{R}}) \geq w_i(\mathcal{R})$ a existuje aspoň jeden prechodný stav, pre ktorý je nerovnosť ostrá. □

2.6.2 Algoritmus value iteration

Ukázali sme, že algoritmus policy iteration skonverguje po konečnom počte iterácií, a tak existuje jednoznačné číslo $\tilde{g}$ a až na aditívnu konštantu jednoznačné čísla $\tilde{v_i}, i \in S$ splňujúce

$$\tilde{v_i} = \max_{r_i \in K(i)} \left\{ c_i(r_i) - \tilde{g} + \sum_{j \in S} p_{ij}(r_i)\tilde{v_j} \right\}, i \in S. \tag{2.36}$$

Stačí si uvedomiť, že v poslednej a predposlednej iterácii sa vygeneruje rovnaké riadenie $\tilde{\mathcal{R}}$. V poslednom kroku algoritmu teda riadenie $\tilde{\mathcal{R}}$ maximalizuje výraz (2.32), a tak podľa vety 2.10 platí (2.36), pričom je $\tilde{g} = g(\tilde{\mathcal{R}})$ a $v_i = v_i(\tilde{\mathcal{R}})$ v zmysle kroku 2 Policy iteration algoritmu. Ďalej je zrejmé, že každé riadenie $\mathcal{R}$, ktoré maximalizuje (2.36) je optimálne vhľadom na priemerný výnos, pričom maximálny priemerný výnos je jednoznačne daný konštantou $\tilde{g}$.

Predpokladajme, že je $c_i(r_i) > 0, \forall i \in S, r_i \in K(i)$ a že pre všetky prípustné stacionárne riadenia je reťazec aperiodický a má len jednu triedu trvalých stavov. Value iteration algoritmus potom môžeme zostaviť nasledovne.

- Krok 0 - Inicializácia
  Položíme $u_i(0) = 0,\ i \in S$. Ďalej nastavíme $n := 1$ a postúpime na samotný iteračný algoritmus.

- Krok 1 - Value iteration
  Pre každý stav $i \in S$ spočítame

  $$u_i(n) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} p_{ij}(r_i)u_j(n-1) \right\}. \tag{2.37}$$

  Nech $\mathcal{R}$ je stacionárne riadenie, ktorého príslušné rozhodnutia $R_i,\ i \in S$ maximalizujú pravú stranu výrazu.

- Krok 2 - Hranice
  Spočítame hranice

  $$L_n = \min_{i \in S}\{u_i(n) - u_i(n-1)\}, \quad U_n = \max_{i \in S}\{u_i(n) - u_i(n-1)\}.$$

- Krok 3 - Ukončovacie pravidlo
  Ak

  $$0 \leq U_n - L_n \leq \epsilon L_n$$

  kde $\epsilon > 0$ je predom určený koeficient presnosti, algoritmus skončí a za optimálne riadenie vezmeme $\mathcal{R}$. Volíme napríklad $\epsilon = 10^{-3}$. V opačnom prípade položíme $n := n + 1$ a pokračujeme krokom 1.

Ak algoritmus skončí po n iteráciách, tak priemerný výnos vygenerovaného riadenia $\mathcal{R}_n$ sa nemôže od teoretického maximálneho priemerného výnosu vychyľovať o viac ako $100\epsilon\%$. Podľa nasledujúcej vety 2.15 totiž platí

$$0 \leq \frac{\tilde{g} - g(\mathcal{R}_n)}{\tilde{g}} \leq \frac{U_n - L_n}{L_n} \leq \epsilon$$

Veta 2.15 *Nech pre každé stacionárne riadenie $\mathcal{R}$ má príslušný Markovov reťazec len jedinú triedu trvalých stavov. Nech $\mathcal{R}_n$ značí riadenie, ktoré bolo vygenerované v n-tom iteračnom cykle algoritmu. Potom platí*

$$L_n \leq g(\mathcal{R}_n) \leq \tilde{g} \leq U_n, \tag{2.38}$$

kde $\tilde{g}$ značí maximálny očakávaný dlhodobý výnos za časovú jednotku.

Dôkaz: Zvoľme ľubovolné stacionárne riadenie $\mathcal{R}$. Podľa kroku 1 Value iteration algoritmu je

$$u_i(n) \geq c_i(R_i) + \sum_{j \in S} p_{ij}(R_i) u_j(n-1),\ i \in S \tag{2.39}$$

Ďalej z definície hornej medze je $U_n \geq u_i(n) - u_i(n-1)$ pre $\forall i \in S$, a tak s využitím (2.39) platí $U_n + u_i(n-1) \geq c_i(R_i) + \sum_{j \in S} p_{ij}(R_i) u_j(n-1)$ pre $\forall i \in S$. Môžeme teda písať

$$u_i(n-1) \geq c_i(R_i) - U_n + \sum_{j \in S} p_{ij}(R_i) u_j(n-1),\ i \in S \tag{2.40}$$

Podľa vety 2.8 je $U_n \geq g(\mathcal{R})$. Keďže $\mathcal{R}$ bolo zvolené ľubovolne, platí

$$U_n \geq \max_{\mathcal{R}} g(\mathcal{R}) = \tilde{g},$$

čo dokazuje prvú nerovnosť. Ďalej pre riadenie $\mathcal{R}_n$ platí

$$u_i(n) = c_i(\mathcal{R}_n) + \sum_{j \in S} p_{ij}(\mathcal{R}_n) u_j(n-1),\ i \in S \tag{2.41}$$

Z definície dolnej medze máme $u_i(n-1) \leq u_i(n) - L_n$, pre $\forall i \in S$, a tak s využitím (2.41) dostávame

$$u_i(n-1) \leq c_i(\mathcal{R}_n) - L_n + \sum_{j \in S} p_{ij}(\mathcal{R}_n) u_j(n-1),\ i \in S.$$

Použijeme vetu 2.8 podľa ktorej platí $L_n \leq g(\mathcal{R}_n)$. Tým sme ukázali platnosť celého tvrdenia. □

Poznámka 2.16 *Ak v* (2.39) *položíme* $n = k+1$, $\mathcal{R} = \mathcal{R}_k$ *a v* (2.41) $n = k$, *odčítaním dostaneme*

$$u_i(k+1) - u_i(k) \geq \sum_{j \in S} p_{ij}(\mathcal{R}_k)[u_j(k) - u_j(k-1)] \geq L_k,\ i \in S, \tag{2.42}$$

takže máme $L_{k+1} \geq L_k$. *Podobne ak položíme v* (2.41) $n = k+1$ *a v* (2.39) $n = k$, $\mathcal{R} = \mathcal{R}_{k+1}$ *dostaneme nakoniec* $U_{k+1} \leq U_k$. *Horná hranica teda tvorí nerastúcu a dolná neklesajúcu postupnosť.*

Poznámka 2.17 *Výsledky z predošlej vety môžeme použiť k vylepšeniu policy iteration algoritmu. V* n*-tej iterácii spočítajme*

$$L_n = \min_{i \in S} \max_{r_i \in K(i)} \{c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R}) - v_i(\mathcal{R})\}$$

$$U_n = \max_{i \in S} \max_{r_i \in K(i)} \{c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R}) - v_i(\mathcal{R}),\}$$

Ak vo vete 2.15 položíme $u_i(n-1) = v_i(\mathcal{R}),\ i \in S$ *a máme* $u_i(n) = \max_{r_i \in K(i)}\{c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) v_j(\mathcal{R})\},\ i \in S$ *dostaneme opäť výsledok* $L_n \leq \tilde{g} \leq U_n$.

Konvergenciu value iteration algoritmu za nami definovaných podmienok zaručuje veta 2.19. Vyslovme najprv nasledujúce pomocné tvrdenie.

Lemma 2.18 *Nech* $\{\mathbf{P}(k), k = 1, 2, \ldots, M\}$ *je konečná množina matíc s kladnými prvkami so spektrálnym polomerom menším než 1, ktoré splňujú podmienku dynamického programovania (v množine sú obsiahnuté všetky matice, ktoré vzniknú nezávislým poskladaním prípustých riadkov). Potom existuje striktne pozitívny vektor* $\mathbf{u}$ *a kladné číslo* $\rho < 1$ *také že* $\mathbf{P}(k)\mathbf{u} \leq \rho\mathbf{u}$ *pre* $\forall k = 1, 2, \ldots, M$.

Dôkaz: Označme pre $k = 1, 2, \ldots, M$ symbolom $\rho^{(k)}$ v absolútnej hodnote najväčšie vlastné číslo matice $\mathbf{P}(k)$ a $\mathbf{u}^{(k)}$ príslušný pravý vlastný vektor. Za platnosti podmienky dynamického programovania je možné preusporiadať matice $\mathbf{P}(k)$ tak, aby platilo $[\mathbf{P}(k) - \mathbf{P}(k-1)]\mathbf{u}^{(k-1)} > 0,\ k = 2, 3, \ldots, M$. Pre $k = 2, 3, \ldots, M$ máme $\mathbf{P}(k-1)\mathbf{u}^{(k-1)} = \rho^{(k-1)}\mathbf{u}^{(k-1)}$ a $\mathbf{P}(k)\mathbf{u}^{(k)} = \rho^{(k)}\mathbf{u}^{(k)}$, a tak môžeme postupne odvodiť

$$\mathbf{P}(k)\mathbf{u}^{(k)} - \mathbf{P}(k-1)\mathbf{u}^{(k-1)} = \rho^{(k)}\mathbf{u}^{(k)} - \rho^{(k-1)}\mathbf{u}^{(k-1)}$$

$$\begin{aligned} & \mathbf{P}(k)\mathbf{u}^{(k)} + \mathbf{P}(k)\mathbf{u}^{(k-1)} - \mathbf{P}(k)\mathbf{u}^{(k-1)} - \mathbf{P}(k-1)\mathbf{u}^{(k-1)} = \\ = \ & \rho^{(k)}\mathbf{u}^{(k)} + \rho^{(k)}\mathbf{u}^{(k-1)} - \rho^{(k)}\mathbf{u}^{(k-1)} - \rho^{(k-1)}\mathbf{u}^{(k-1)} \end{aligned}$$

$$\mathbf{P}(k)[\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}] + [\mathbf{P}(k) - \mathbf{P}(k-1)]\mathbf{u}^{(k-1)} = \rho^{(k)}[\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}] + [\rho^{(k)} - \rho^{(k-1)}]\mathbf{u}^{(k-1)}$$

Označme $\mathbf{v}^{(k)}$ ľavý vlastný vektor prislúchajúci $\rho^{(k)}$, takže $\mathbf{v}^{(k)}\mathbf{P}(k) = \rho^{(k)}\mathbf{v}^{(k)},\ k = 1, 2, \ldots, M$. Prenásobením oboch strán predošlej rovnoti zľava vektorom $\mathbf{v}^{(k)}$ nakoniec dostaneme

$$\mathbf{v}^{(k)}[\mathbf{P}(k) - \mathbf{P}(k-1)]\mathbf{u}^{(k-1)} = [\rho^{(k)} - \rho^{(k-1)}]\mathbf{v}^{(k)}\mathbf{u}^{(k-1)}$$

z čoho plynie $\rho^{(k)} > \rho^{(k-1)},\ k = 2, 3, \ldots, M$. Položíme teda $\rho = \rho^{(M)} < 1$ a $\mathbf{u} = \mathbf{u}^{(M)}$, čím dostaneme tvrdenie vety. □

Veta 2.19 *Nech pre každé stacionárne riadenie $\mathcal{R}$ má príslušný Markovov reťazec len jedinú triedu trvalých stavov, ktoré su navyše aperiodické. Potom horná a dolná medz generovaná value iteration algoritmom konverguje k maximálnemu očakávanému dlhodobému priemernému výnosu, t.j. platí*

$$\lim_{n\to\infty} L_n = \lim_{n\to\infty} U_n = \tilde{g}.$$

Dôkaz: Označme opäť $\mathcal{R}_n$ riadenie vygenerované algoritmom v n-tej iterácii. S využitím vektorového značenia teda platí

$$\mathbf{u}(n+1) = \max_{\mathcal{R}}[\mathbf{c}(\mathcal{R}) + \mathbf{P}(\mathcal{R})\mathbf{u}(n)] = \mathbf{c}(\mathcal{R}_n) + \mathbf{P}(\mathcal{R}_n)\mathbf{u}(n) \tag{2.43}$$

Zvoľme ľubovolné riadenie optimálne vzhľadom k dlhodobému priemernému výnosu a označme ho $\tilde{\mathcal{R}}$.

V prvom kroku dôkazu predpokladajme, že za tohto riadenia je príslušný Markovov reťazec nerozložitelný, teda uvažujeme len trvalé nenulové stavy. Poznamenajme ale, že matice $\mathbf{P}(\mathcal{R}_n)$ nerozložitelné byť nemusia. Vzorec (2.36) potom môžeme prepísať do vektorového tvaru

$$\tilde{\mathbf{v}} + \tilde{\mathbf{g}} = \max_{\mathcal{R}}[\mathbf{c}(\mathcal{R}) + \mathbf{P}(\mathcal{R})\tilde{\mathbf{v}}] = \mathbf{c}(\tilde{\mathcal{R}}) + \mathbf{P}(\tilde{\mathcal{R}})\tilde{\mathbf{v}}. \tag{2.44}$$

Pre $n \in N$ definujme $\mathbf{y}(n) = \mathbf{u}(n) - [n\tilde{\mathbf{g}} + \tilde{\mathbf{v}}]$. Podľa (2.43) a (2.44) je

$$\begin{aligned} \mathbf{y}(n+1) &= \mathbf{u}(n+1) - [(n+1)\tilde{\mathbf{g}} + \tilde{\mathbf{v}}] \\ &= \mathbf{c}(\mathcal{R}_n) + \mathbf{P}(\mathcal{R}_n)\mathbf{u}(n) - n\tilde{\mathbf{g}} - [\mathbf{c}(\tilde{\mathcal{R}}) + \mathbf{P}(\tilde{\mathcal{R}})\tilde{\mathbf{v}}] \end{aligned}$$

Ďalej zrejme platí

$$\begin{aligned} \mathbf{y}(n+1) &\leq \mathbf{c}(\mathcal{R}_n) + \mathbf{P}(\mathcal{R}_n)\mathbf{u}(n) - n\tilde{\mathbf{g}} - [\mathbf{c}(\mathcal{R}_n) + \mathbf{P}(\mathcal{R}_n)\tilde{\mathbf{v}}] \\ &= \mathbf{P}(\mathcal{R}_n)[\mathbf{u}(n) - \tilde{\mathbf{v}} - n\tilde{\mathbf{g}}] = \mathbf{P}(\mathcal{R}_n)\mathbf{y}(n) \end{aligned}$$

$$\begin{aligned} \mathbf{y}(n+1) &\geq \mathbf{c}(\tilde{\mathcal{R}}) + \mathbf{P}(\tilde{\mathcal{R}})\mathbf{u}(n) - n\tilde{\mathbf{g}} - [\mathbf{c}(\tilde{\mathcal{R}}) + \mathbf{P}(\tilde{\mathcal{R}})\tilde{\mathbf{v}}] \\ &= \mathbf{P}(\tilde{\mathcal{R}})[\mathbf{u}(n) - \tilde{\mathbf{v}} - n\tilde{\mathbf{g}}] = \mathbf{P}(\tilde{\mathcal{R}})\mathbf{y}(n). \end{aligned}$$

Celkovo teda dostávame

$$\mathbf{P}(\tilde{\mathcal{R}})\mathbf{y}(n) \leq \mathbf{y}(n+1) \leq \mathbf{P}(\mathcal{R}_n)\mathbf{y}(n). \tag{2.45}$$

Nakoniec z (2.45) dostaneme, že pre $\forall i \in S$ platí

$$\min_{j\in S} y_j(n) \leq y_i(n+1) \leq \max_{j\in S} y_j(n).$$

Označme $\check{z}(n) = \min_{j\in S} y_j(n)$ a $\hat{z}(n) = \max_{j\in S} y_j(n)$. L'akho sa podľa predchádzajúceho vzorca presvedčíme, že postupnosť $\{\check{z}(n)\}_{n=1}^{\infty}$ je neklesajúca, postupnosť $\{\hat{z}(n)\}_{n=1}^{\infty}$ je nerastúca a obe sú ohraničené, z čoho vyplýva že existujú ich konečné limity. Označne teda $\check{z} = \lim_{n\to\infty} \check{z}(n)$ a $\hat{z} = \lim_{n\to\infty} \hat{z}(n)$. Zrejme platí, že $\check{z} \leq \hat{z}$. Naším cielom bude ukázať, že je dokonca $\check{z} = \hat{z}$. Predpokladajme pre spor, že $\check{z} < \hat{z}$.

Množina uvažovaných stavov S je konečná, a tak existuje stav $a \in S$ a rastúca podpostupnosť indexov $m_k \to \infty$ taká, že $y_a(m_k) = \hat{z}(m_k)$. Ďalej existuje stav $b \in S$ a rastúca postupnosť indexov $n_k \to \infty$ taká, že $y_b(n_k + m_k) = \check{z}(n_k + m_k)$. Platí teda

$$\lim_{k\to\infty} y_a(m_k) = \hat{z}, \qquad \lim_{k\to\infty} y_b(n_k + m_k) = \check{z}. \tag{2.46}$$

Iteráciou ľavej časti vzorca (2.45) dostaneme

$$\mathbf{y}(n+m) \geq (\mathbf{P}(\tilde{\mathcal{R}}))^n \mathbf{y}(m). \tag{2.47}$$

Matica $\mathbf{P}(\tilde{\mathcal{R}})$ je za nami vyslovených predpokladov ergodická (nerozložitelná aperiodická stochastická matica), a tak platí

$$\lim_{n\to\infty} (\mathbf{P}(\tilde{\mathcal{R}}))^n = \mathbf{P}^*(\tilde{\mathcal{R}}) = \begin{pmatrix} \Pi_1 & \Pi_2 & \dots & \Pi_N \\ \Pi_1 & \Pi_2 & \dots & \Pi_N \\ \vdots & \vdots & \vdots & \vdots \\ \Pi_1 & \Pi_2 & \dots & \Pi_N \end{pmatrix},$$

pričom pre $\forall i \in S$ je $\Pi_i > 0$ a $\sum_{i\in S} \Pi_i = 1$. Pre b-ty riadok rovnice (2.47) dostaneme

$$\begin{aligned} y_b(n_k + m_k) &\geq \sum_{l\in S} [(\mathbf{P}(\tilde{\mathcal{R}}))^{n_k}]_{bl}\, y_l(m_k) \\ &\geq \sum_{l\in S, l\neq a} [(\mathbf{P}(\tilde{\mathcal{R}}))^{n_k}]_{bl}\, \check{z}(m_k) + [(\mathbf{P}(\tilde{\mathcal{R}}))^{n_k}]_{ba}\, y_a(m_k) \end{aligned}$$

Limitným prechodom pre k $\to \infty$ dostaneme

$$\check{z} \geq \sum_{l\neq a} \Pi_l \check{z} + \Pi_a \hat{z},$$

čo je spor s predpokladom $\check{z} < \hat{z}$. Existuje teda $\lim_{n\to\infty} \mathbf{y}(n) = \mathbf{y}$, kde $\mathbf{y} \geq \mathbf{P}^*(\tilde{\mathcal{R}})\mathbf{y}$. Pretože $\mathbf{P}^*(\tilde{\mathcal{R}})$ je stochastická matica, musí byť $\mathbf{y}$ konštantný vektor. Pripomeňme, že

$\mathbf{y} = \mathbf{u}(n) - [n\tilde{\mathbf{g}} + \tilde{\mathbf{v}}]$ kde $\tilde{\mathbf{v}}$ je vektor určený jednoznačne až na aditívnu konštantu. Ak položíme $\tilde{\mathbf{v}} := \tilde{\mathbf{v}} + \mathbf{y}$ dostaneme $\lim_{n\to\infty} \mathbf{y}(n) = \mathbf{0}$.

Ďalej predpokladajme, že pre každé stacionárne riadenie má reťazec len jedinú triedu trvalých stavov. Z predchádzajúcej časti dôkazu vieme, že pri vhodnej voľbe riešenia $\tilde{\mathbf{v}}$, máme pre stav $i \in S$, ktorý je trvalý aspoň pre jedno optimálne riadenie $\lim_{n\to\infty} y_i(n) = 0$. Pre každé stacionárne riadenie $\mathcal{R}$ zaveďme nasledujúce rozdelenie matice pravdepodobnotí prechodu.

$$\mathbf{P}(\mathcal{R}) = \begin{pmatrix} \mathbf{P}_{TT}(\mathcal{R}) & \mathbf{P}_{TR}(\mathcal{R}) \\ \mathbf{R}_{RT}(\mathcal{R}) & \mathbf{P}_{RR}(\mathcal{R}) \end{pmatrix},$$

kde $\mathbf{P}_{RR}(\mathcal{R})$ je submatica, ktorá odpovedá stavom, ktoré sú trvalé aspoň pre jedno optimálne riadenie. Stavy sú teda prečíslované tak, aby odpovedali predošlému značeniu. Zvoľme ľubovolbé $n \in N$. Pre riadenie $\mathcal{R}_n$ vygenerované v n-tom kroku budeme podľa prvej strany (2.45) písať

$$\begin{pmatrix} \mathbf{y}_T(n+1) \\ \mathbf{y}_R(n+1) \end{pmatrix} \leq \begin{pmatrix} \mathbf{P}_{TT}(\mathcal{R}_n) & \mathbf{P}_{TR}(\mathcal{R}_n) \\ \mathbf{P}_{RT}(\mathcal{R}_n) & \mathbf{P}_{RR}(\mathcal{R}_n) \end{pmatrix} \begin{pmatrix} \mathbf{y}_T(n) \\ \mathbf{y}_R(n) \end{pmatrix}.$$

Odtiaľ pre $\mathbf{y}_T(n+1)$ dostaneme

$$\mathbf{y}_T(n+1) \leq \mathbf{P}_{TT}(\mathcal{R}_n)\mathbf{y}_T(n) + \mathbf{P}_{TR}(\mathcal{R}_n)\mathbf{y}_R(n)$$

a iterovaním tejto nerovnosti nakoniec

$$\begin{aligned} \mathbf{y}_T(n+m) \quad \leq \quad & \mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\mathbf{P}_{TT}(\mathcal{R}_{n+m-2})\ldots\mathbf{P}_{TT}(\mathcal{R}_n)\mathbf{y}_T(n) + \\ & + \mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\ldots\mathbf{P}_{TT}(\mathcal{R}_{n+1})\mathbf{P}_{TR}(\mathcal{R}_n)\mathbf{y}_R(n) + \\ & + \mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\ldots\mathbf{P}_{TT}(\mathcal{R}_{n+2})\mathbf{P}_{TR}(\mathcal{R}_{n+1})\mathbf{y}_R(n+1) + \\ & + \mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\ldots\mathbf{P}_{TT}(\mathcal{R}_{n+3})\mathbf{P}_{TR}(\mathcal{R}_{n+2})\mathbf{y}_R(n+2) + \\ & \ldots \\ & + \mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\mathbf{y}_R(n+m-1). \end{aligned} \tag{2.48}$$

Z predchádzajúcej časti dôkazu vieme, že $\mathbf{y}_R(n) \to 0$ a tak môžeme pre dosť veľké n učiniť $\mathbf{y}_R(n)$ ľubovolne malé. Pre zvolené $n \in N$ teda $\exists\delta_n > 0$ také, že $\forall n' \geq n$ je $\mathbf{y}_R(n') < \delta_n\mathbf{e}$. Zaoberajme sa teda maticovými súčinmi matíc $\mathbf{P}_{TT}$. Pre zvolené $n, m \in N$ uvažujme množinu m matíc

$$M_{n,m} = \{\mathbf{P}_{TT}(\mathcal{R}_{n+m-1}), \mathbf{P}_{TT}(\mathcal{R}_{n+m-2}), \ldots, \mathbf{P}_{TT}(\mathcal{R}_n)\}$$

Nech $\rho_{n,m} < 1$ značí najväčší spektrálny polomer spomedzi všetkých spektrálnych polomerov matíc z $M_{n,m}$. Zvoľme $\epsilon > 0$ a položme $\mathbf{P}^{\epsilon}_{TT} = \mathbf{P}_{TT} + \epsilon\mathbf{E}$, pre $\forall P_{TT} \in M_{n,m}$. Označme $M^{\epsilon}_{n,m} = \{\mathbf{P}^{\epsilon}_{TT}(\mathcal{R}_{n+m-1}), \mathbf{P}^{\epsilon}_{TT}(\mathcal{R}_{n+m-2}), \ldots, \mathbf{P}^{\epsilon}_{TT}(\mathcal{R}_n)\}$ a $\rho^{\epsilon}_{n,m} < 1$ najväčší spektrálny polomer. Množina matíc $M^{\epsilon}_{n,m}$ pre ϵ dostatočne malé splňuje predpoklady lemmy 2.18, a tak platí

$$\mathbf{P}_{TT}(\mathcal{R}_{n+m-1})\mathbf{P}_{TT}(\mathcal{R}_{n+m-2})\ldots\mathbf{P}_{TT}(\mathcal{R}_n)\mathbf{u}^{\epsilon}_{n,m} \leq (\rho^{\epsilon}_{n,m})^m\mathbf{u}^{\epsilon}_{n,m},$$

kde $\mathbf{u}^{\epsilon}_{n,m}$ je striktne pozitívny vlastný vektor prislúchajúci $\rho^{\epsilon}_{n,m}$.

Každý prvok tohoto maticového súčinu je teda menší ako $f(\rho^{\epsilon}_{n,m})^m$ kde $f > 0$ je nejaká vhodne zvolená konštanta. Celkovo teda môžeme nerovnosť (2.48) upraviť na tvar

$$\mathbf{y}_T(n+m) \leq (\rho^{\epsilon}_{m,n})^m f \mathbf{y}_T(n) + \sum_{k=1}^{m} (\rho^{\epsilon}_{m,n})^k f \delta_n,$$

takže $\mathbf{y}_T(n+m) \to 0$ pre $m \to \infty$.

Máme teda

$$L_n = \min_{i \in S}\{u_i(n) - u_i(n-1)\} = \min_{i \in S}\{y_i(n) - y_i(n-1) + \tilde{g}\},$$

a tak platí $\lim_{n\to\infty} L_n = \tilde{g}$ a podobne $\lim_{n\to\infty} U_n = \tilde{g}$ □

Dôsledkom predchádzajúcich viet je že $g(\mathcal{R}_n) \to \tilde{g}$

Poznámka 2.20 *Z predošlého dôkazu navyše plynie, že hodnoty $u_i(n)$ rastú do nekonečna. Z výpočtového hľadiska sa preto môže hodiť vykonať nasledujúcu korekciu algoritmu, ktorá zaručí že iterované hodnoty budú obmedzené. Položíme $h_i(n) = u_i(n) - u_N(n),\ i \in S$ a $k_N(n+1) = u_N(n+1) - u_N(n)$. Z iteračného vzorca (2.37) postupne dostaneme*

$$u_i(n+1) - u_N(n) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} p_{ij}(r_i)[u_j(n) - u_N(n)] \right\},\ i \in S,$$

$$h_i(n+1) + k_N(n+1) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} p_{ij}(r_i) h_j(n) \right\},\ i \in S, \qquad (2.49)$$

Dosadením $h_i(n), i \in S$ do pravej strany (2.49) napočítame nové hodnoty. Keďže pre $\forall n \in N$ platí $h_N(n) = 0$ najprv spočítame

$$k_N(n+1) = \max_{r_N \in K(N)} \left\{ c_N(r_N) + \sum_{j \in S} p_{Nj}(r_N) h_j(n) \right\}, \qquad (2.50)$$

a potom odčítaním tejto hodnoty dopočítame hodnoty $h_i(n+1)$ pre zvyšné stavy. Medze potom napočítame podľa

$$L_n = \min_{i \in S}\{h_i(n+1) + k_N(n+1) - h_i(n)\}, \quad U_n = \max_{i \in S}\{h_i(n+1) + k_N(n+1) - h_i(n)\}.$$

2.6.3 Problém výrobcu hračiek

Nasledujúci problém je rozšírením úlohy (viac stavov a stretégií) z knihy [1]. Výrobca hračiek na konci každého výrobného obdobia hodnotí predajnosť svojej hračky jedným s 5 stavov a to konkrétne

- Stav 1: hračka je veľmi úspešná
- Stav 2: hračka je úspešná
- Stav 3: hračka je priemerná
- Stav 4: hračka je neúspešná
- Stav 5: hračka je veľmi neúspešná

Výrobca má potom možnosť zvoliť určitú stratégiu ďalšieho predaja hračky. Celkovo uvažuje nasledujúce stratégie.

- Akcia 1: predávať bez zmien.
- Akcia 2: investovať do marketingu
- Akcia 3: modernizovať hračku
- Akcia 4: zvýšiť ceny

V prípade voľby akcie 1 sa hračka predáva bez investovania do reklamnej kampane či modernizácie za štandardné ceny. Akcia 2 znamená investovať do reklamnej kampane, čo v určitých situáciách znamená zvýšenie pravdepodobnosti prechodu do lepšieho stavu za cenu nižšieho zisku. Akcia 3 znamená investovanie do inovácie hračky, čo v určitých situáciách podobne ako v prípade akcie 2 spôsobí zvýšenie pravdepodobnosti prechodu do lepšieho stavu za cenu nižšieho zisku. Akcie 1,2 a 3 sú prípustné pre akékoľvek stavy. Akcia zvýšiť ceny naopak spôsobí zvýšenie pravdepodobnosti prechodu do horších stavov, avšak zvýši sa možný zisk. Výrobca má skúsenosť, že túto akciu stačí uvažovať, ak hračka nieje veľmi neuspešná alebo neúspešná. Výrobca má s predajom tejto hračky dlhodobé skúsenosti, na základe ktorých boli odhadnuté pravdepodobnosti prechodu pre jednotlivé stavy. Z každým prechodom je spojený nejaký výnos z predaja od ktorého sú odčítané náklady na danú stratégiu. Pravdepodobnosti prechodu a výnosy po 1 kroku sa nachádzajú v tabuľke 2.11. Výrobca má ešte fixné náklady v hodnote 4 pre akýkoľvek prechod a zvolenú stratégiu. Táto hodnota nebola do výpočtu zahrnutá, keďže oplyvní výnosy po 1 prechode stále rovnakým dielom.

i	r	$p_{ij}(r)$					$c_i(r)$
		$j =$ 1	2	3	4	5	
1	1	0.2	0.5	0.15	0.1	0.05	10.2
1	2	0.25	0.55	0.1	0.075	0.025	8.525
1	3	0.2	0.6	0.1	0.05	0.05	7.9
1	4	0.15	0.45	0.2	0.125	0.075	10.95
2	1	0.2	0.35	0.25	0.15	0.05	8.75
2	2	0.35	0.35	0.15	0.1	0.05	7.55
2	3	0.2	0.45	0.2	0.1	0.05	7.75
2	4	0.1	0.2	0.4	0.2	0.1	9.3
3	1	0.1	0.25	0.2	0.3	0.15	8.0
3	2	0.15	0.35	0.25	0.2	0.05	7.775
3	3	0.2	0.35	0.3	0.1	0.05	7.163
3	4	0.025	0.15	0.2	0.4	0.225	6.712
4	1	0.1	0.15	0.2	0.35	0.2	4.25
4	2	0.15	0.25	0.25	0.2	0.15	2.725
4	3	0.2	0.25	0.3	0.15	0.1	4.0125
5	1	0.05	0.15	0.2	0.3	0.3	1.25
5	2	0.05	0.2	0.3	0.25	0.2	0.175
5	3	0.15	0.2	0.3	0.25	0.1	1.213

Tabulka 2.11: Pravdepodobnosti prechodu a očakávaný výnos

Úlohu najprv vyriešime policy iteration algoritmom. Algoritmus sa zastaví už po 3 iteráciách s nasledujúcim optimálnym riadením.

- Ak je hračka veľmi úspešná (stav1) zvolí sa stratégia zvýšenia cien.
- Ak je hračka úspešná (stav2) zvolí sa stratégia predávať bez zmien.
- Ak je hračka priemerná (stav3) zvolí sa stratégia investovať do marketingu.
- Ak je hračka neúspešná (stav4) zvolí sa stratégia investovať do inovácie.
- Ak je hračka veľmi neúspešná (stav5) zvolí sa stratégia investovať do inovácie.

Výsledky sústavy ako aj výstupné riadenie pre každú iteráciu sú uvedené v tabuľkách 2.12 a 2.13.

Iter.	Vstupné riadenie	Riešenie sústavy rovníc					
		w_1	w_2	w_3	w_4	w_5	g
1	$\{1,1,1,1,1\}$	13.9659	11.8030	8.6286	3.8929	0.0000	6.612
2	$\{1,1,3,3,3\}$	10.5082	8.6567	7.2614	3.4166	0.0000	7.553
3	$\{4,1,2,3,3\}$	10.7429	8.6655	7.3247	3.4277	0.0000	7.611

Tabulka 2.12: Riešenie - policy iteration

Iter.	Vypočítané riadenie				
	stav= 1	2	3	4	5
0	1	1	1	1	1
1	1	1	3	3	3
2	4	1	2	3	3
3	4	1	2	3	3

Tabulka 2.13: Riešenie - policy iteration

Pomocou algoritmu value iteration s nastavením $\epsilon = 0.001$ taktiež dospejeme k rovnakému riadeniu, pričom algoritmus sa zastaví po 5 iteráciach. Výsledky z jednotlivých iterácií sa nachádzajú v tabuľke 2.14.

It. (n)	Vypočítané hodnoty					Riadenie	L_n	U_n
	$u_1(n)$	$u_2(n)$	$u_3(n)$	$u_4(n)$	$u_5(n)$			
1	10.95	9.30	8.00	4.25	1.25	$\{4,4,1,1,1\}$	1.25	10.95
2	19.003	16.895	15.585	11.690	8.303	$\{4,1,2,3,3\}$	7.053	8.053
3	26.604	24.529	23.188	19.296	15.870	$\{4,1,2,3,3\}$	7.568	7.634
4	34.218	32.141	30.800	26.903	23.476	$\{4,1,2,3,3\}$	7.606	7.614
5	41.830	39.752	38.412	34.515	31.087	$\{4,1,2,3,3\}$	7.611	7.612

Tabulka 2.14: Riešenie - value iteration

2.6.4 Problém cestujúceho opravára

Opravár cestuje každý týždeň k zákazníkom z piatich veľkých miest, pričom sa riadi presne daným rozvrhom. V pondelok sa nachádza v Amsterdame (mesto 1), v utorok v Rotterdame (mesto 2), v stredu v Bruseli (mesto 3), vo štvrtok v Aachene (mesto 4) a v piatok v Arnheme (mesto 5). Úlohou opravára je skontrolovať funkčnosť istého dôležitého elementu v elektrickom zariadení, ktoré prenajíma firma zamestnávajúca opravára. Po kontrole je niekedy nutné element vymeniť. Pravdepodobnostné rozdelenie počtu potrebných výmen v meste j je dané pravdepodobnosťami $\{p_j(k),\ k = 0, 1, 2, \ldots K\}$, pre každé $j = 1, \ldots 5$. Počty potrebných výmen v posebe nasledujúcich dňoch sú navzájom nezávislé. Opravár je schopný uniesť najviac M náhradných dielov. Ak počet náhradných dielov, ktoré mal pri sebe opravár na začiatku pracovného dňa nestačilo na pokrytie dopytu v danom meste, je firma povinná vyslať nasledujúci deň iného opravára, ktorý dokončí ostávajúce výmeny. Cena takejto špeciálnej cesty do mesta j je rovná K_j. Na konci každého dňa sa opravár môže rozhodnúť, či si nechá doplniť zásobu náhradných dielov. Ak sa v meste j rozhodne pre doplnenie, dorazí mu ráno zásielka v cene a_j a cestuje do ďalšieho mesta. Opravár je rezident mesta 5. Úlohou je nájsť optimálne riadenie (rozhodnutia opravára v jednotlyvých situáciách) tak, aby sme minimalizovali priemerný očakávaný náklad. Úlohu vyriešime policy aj value iteration algoritmom pre numerické data $M = 20,\ K = 9,\ K_j = 200,\ j = 1, \ldots, 5,\ a_1 = 60,\ a_2 = 30,\ a_3 = 50,\ a_4 = 25,\ a_5 = 100$ a pravdepodobnosti $p_j(k),\ j = 1, \ldots, 5,\ k = 0, \ldots, 9$ dané tabuľkou 2.15.

Mesto	Počet opráv									
	0	1	2	3	4	5	6	7	8	9
1	0.075	0.225	0.3	0.234	0.117	0.039	0.009	0.001	0	0
2	0.04	0.156	0.267	0.267	0.172	0.073	0.021	0.004	0	0
3	0.01	0.06	0.161	0.251	0.251	0.167	0.074	0.021	0.005	0
4	0.005	0.034	0.111	0.212	0.26	0.213	0.116	0.041	0.007	0.001
5	0.002	0.018	0.07	0.165	0.245	0.247	0.162	0.07	0.018	0.003

Tabulka 2.15: Pravdepodnosti spotreby dielov

Ako prvé musíme zadanie transformovať na tvar, ktorý zodpovedá úlohe stochastického dynamického programovania. Uvedomme si, že opravár sa na konci dňa v každom meste môže ocitnúť v $M + 2$ situáciách. Prvá situácia je tá, že opravár nemal dostatok náhradných dielov a tak musí byť ďalší deň vyslaný špeciálny opravár (túto sitáciu, z ktorej plynú penalizačné náklady, budeme značiť písmenom P). Ďalšie situácie vystihujú stav opravárových zásob na konci dňa. Celkovo teda definujeme $(M + 2) * 5$ stavov. Opravár môže uskutočniť 2 akcie, a to "požiadať o doplnenie zásob"(značíme písmenom D) alebo "nepožiadať o doplnenie zásob"(značíme písmenom ND) v každom stave okrem stavu, v ktorom má zásobu M (vtedy "požiadať o doplnenie zásob"nemá zrejme zmysel) alebo penalizačnom stave (vtedy zrejme nemá zmysel nepožiadať o doplnenie zásob). Definícia stavov je popísaná v nasledujúcej

tabuľke 2.16.

Stav	Mesto	Stav zásob	Akcie
1	1	P	Len doplnenie
2	1	0	Obe
3	1	1	Obe
4	1	2	Obe
5	1	3	Obe
⋮	⋮	⋮	⋮
21	1	19	Obe
22	1	20	Len nedopliť
23	2	P	Len doplnenie
24	2	0	Obe
⋮	⋮	⋮	⋮
43	2	19	Obe
44	2	20	Len nedoplniť
45	3	P	Len doplnenie
46	3	0	Obe
⋮	⋮	⋮	⋮
65	3	19	Obe
66	3	20	Len nedoplniť
67	4	P	Len doplnenie
68	4	0	Obe
⋮	⋮	⋮	⋮
87	4	19	Obe
88	4	20	Len nedoplniť
89	5	P	Len doplnenie
90	5	0	Obe
⋮	⋮	⋮	⋮
109	5	19	Obe
110	5	20	Len nedoplniť

Tabulka 2.16: Definícia stavov

Ako ďalšie je potrebné si uvedomiť ako vyzerajú pravdepodobnosti prechodov medzi takýmito stavmi. Zrejme ak sa nachádzame v meste j, môžeme prejsť len do mesta $j + 1$ (respektíve 1 ak je východiskovým mesto 5). Pravdepodobnosti prechodu do iných miest musia byť nutne nulové. V bloku pre mesto, do ktorého opravár odchádza, budú pravdepodobnosti prechodu pre dané razhodnutie dané transformáciou pravdepodonností z tabuľky 2.15. Napríklad ak sa opravár nachádza v meste 5, tak pravdepodobnosti prechodu pre akékoľvek rozhodnutie budú v prípade výstupných stavov 23 až 110 nulové. V prípade zvyšných stavov, koré popisujú stav zásob na konci dňa v meste 1 sú pravdepodobnosti prechodu popísané v nasledujúcej tabuľke 2.17. Situácia je pre ostatné mestá rovnaká. Čitateľ si už ľahko domyslí tvar matice pravdepodobností prechodu. Ak sa opravár rozhodne pre nedoplnenie zásob, je zrejme cena

prechodu 0. V opačnom prípade je cena v meste j daná a_j, prípadne $a_j + K_j$ ak sa dostaneme do penalizačného stavu.

Výstupné mesto 5												
Výstup zásob	P	okrem P	0	1	2	3	4	5	6	7	...	20
Rozhodnutie	D	D	ND	ND	ND	ND	ND	ND	ND	ND	...	ND
Náklad	300	100	0	0	0	0	0	0	0	0	...	0
Vstup do 1:												
PENAL	0	0	0.925	0.7	0.4	0.166	0.01	0.001	0	0	...	0
0	0	0	0.075	0.225	0.3	0.234	0.039	0.009	0.001	0	...	0
1	0	0	0	0.075	0.225	0.3	0.117	0.039	0.009	0.001	...	0
2	0	0	0	0	0.075	0.225	0.234	0.117	0.039	0.009	...	0
3	0	0	0	0	0	0.075	0.3	0.234	0.117	0.039	...	0
4	0	0	0	0	0	0	0.225	0.3	0.234	0.117	...	0
5	0	0	0	0	0	0	0.075	0.225	0.3	0.234	...	0
6	0	0	0	0	0	0	0	0.075	0.225	0.3	...	0
7	0	0	0	0	0	0	0	0	0.075	0.225	...	0
8	0	0	0	0	0	0	0	0	0	0.075	...	0
9	0	0	0	0	0	0	0	0	0	0	...	0
10	0	0	0	0	0	0	0	0	0	0	...	0
11	0	0	0	0	0	0	0	0	0	0	...	0
12	0	0	0	0	0	0	0	0	0	0	...	0
13	0.001	0.001	0	0	0	0	0	0	0	0	...	0.001
14	0.009	0.009	0	0	0	0	0	0	0	0	...	0.009
15	0.039	0.039	0	0	0	0	0	0	0	0	...	0.039
16	0.117	0.117	0	0	0	0	0	0	0	0	...	0.117
17	0.234	0.234	0	0	0	0	0	0	0	0	...	0.234
18	0.3	0.3	0	0	0	0	0	0	0	0	...	0.3
19	0.225	0.225	0	0	0	0	0	0	0	0	...	0.225
20	0.075	0.075	0	0	0	0	0	0	0	0	...	0.075

Tabulka 2.17: Pravdepodobnosti prechodu z mesta 5

Úlohu najprv vyriešime policy iteration algoritmom, ktorý sa zastaví po 6 iteráciách. Na vyriešenie bol použitý program naprogramovaný v programovacom jazyku JAVA, ktorý bol navrhnutý tak aby hľadal riadenie maximalizujúce výnos. Keďže v tejto úlohe minimalizujeme náklad, bol vstup do programu upravený tak, že ohodnotenie bolo vynásobené -1. Výsledná stratégia je nasledujúca:

- Mesto 1: Opravár doplní zásoby ak klesnú pod 4 súčiastky
- Mesto 2: Opravár doplní zásoby ak klesnú pod 9 súčiastok
- Mesto 3: Opravár doplní zásoby ak klesnú pod 5 súčiastok
- Mesto 4: Opravár doplní zásoby ak klesnú pod 4 súčiastky
- Mesto 5: Opravár doplní zásoby ak klesnú pod 3 súčiastky

V tabuľke 2.18 sú uvedené napočítané hodnoty priemerného očakávaného výnosu g pre riadenia vstupujúce do jednotlivých iterácií. Pre úplnosť ešte v tabuľkách 2.19, 2.20 a 2.6.4 uvádzame pre jednotlivé mestá výstupné riadenia z jednotlivých iterácií.

Všimnime si, že napočítané riadenia nemusia mať práve jeden deliaci stav (pre stavy vystihujúce vyššiu kapacitu zásob v danom meste stále rovnaké rozhodnutie a

Iterácia	0	1	2	3	4	5
Priemerný výnos a	-51.695	-10.351	-8.491	-7.905	-7.304	-7.228

Tabulka 2.18: Priemerný výnosu g pre riadenie vstupujúce do danej iterácie

(a) Riadenie pre mesto 1

	Iterácia						
	0	1	2	3	4	5	6
P	D	D	D	D	D	D	D
0	D	D	D	D	D	D	D
1	D	D	D	D	D	D	D
2	D	D	D	D	D	D	D
3	D	ND	ND	D	D	D	D
4	D	ND	ND	ND	ND	ND	ND
5	D	ND	ND	ND	ND	ND	ND
6	D	ND	ND	ND	ND	ND	ND
7	D	ND	ND	ND	ND	ND	ND
8	D	ND	ND	ND	ND	ND	ND
9	D	ND	ND	ND	ND	ND	ND
10	D	ND	ND	ND	ND	ND	ND
11	D	ND	ND	ND	ND	ND	ND
12	D	ND	ND	ND	ND	ND	ND
13	D	ND	ND	ND	ND	ND	ND
14	D	ND	ND	ND	ND	ND	ND
15	D	ND	ND	ND	ND	ND	ND
16	D	ND	ND	ND	ND	ND	ND
17	D	ND	ND	ND	ND	ND	ND
18	D	ND	ND	ND	ND	ND	ND
19	D	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND

(b) Riadenie pre mesto 2

	Iterácia						
	0	1	2	3	4	5	6
P	D	D	D	D	D	D	D
0	D	D	D	D	D	D	D
1	D	D	D	D	D	D	D
2	D	D	D	D	D	D	D
3	D	D	D	D	D	D	D
4	D	D	D	D	D	D	D
5	D	ND	D	D	D	D	D
6	D	ND	D	D	D	D	D
7	D	ND	D	D	D	D	D
8	D	ND	ND	D	D	D	D
9	D	ND	ND	ND	ND	ND	ND
10	D	ND	ND	ND	ND	ND	ND
11	D	ND	ND	ND	ND	ND	ND
12	D	ND	ND	ND	ND	ND	ND
13	D	ND	D	D	ND	ND	ND
14	D	ND	D	D	ND	ND	ND
15	D	ND	D	D	ND	ND	ND
16	D	ND	D	D	ND	ND	ND
17	D	ND	D	ND	ND	ND	ND
18	D	ND	ND	ND	ND	ND	ND
19	D	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND

Tabulka 2.19: Výsledky policy iteration

pre stavy popisujúce rovnakú a nižšiu kapacitu taktiež rovanké ale opačné riadenie) ako je tomu napríklad v prípade mesta 2. Výsledné riadenie pre každé mesto už však toto pravidlo splňuje, čo je predvídatelné.

(a) Riadenie pre mesto 3

	Iterácia						
	0	1	2	3	4	5	6
P	D	D	D	D	D	D	D
0	D	D	D	D	D	D	D
1	D	D	D	D	D	D	D
2	D	D	D	D	D	D	D
3	D	D	D	D	D	D	D
4	D	D	D	D	D	D	D
5	D	ND	D	D	D	ND	ND
6	D	ND	ND	ND	ND	ND	ND
7	D	ND	ND	ND	ND	ND	ND
8	D	ND	ND	ND	ND	ND	ND
9	D	ND	D	ND	ND	ND	ND
10	D	ND	D	ND	ND	ND	ND
11	D	ND	D	ND	ND	ND	ND
12	D	ND	D	ND	ND	ND	ND
13	D	ND	D	ND	ND	ND	ND
14	D	ND	ND	ND	ND	ND	ND
15	D	ND	ND	ND	ND	ND	ND
16	D	ND	ND	ND	ND	ND	ND
17	D	ND	ND	ND	ND	ND	ND
18	D	ND	ND	ND	ND	ND	ND
19	D	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND

(b) Riadenie pre mesto 4

	Iterácia						
	0	1	2	3	4	5	6
P	D	D	D	D	D	D	D
0	D	D	D	D	D	D	D
1	D	D	D	D	D	D	D
2	D	D	D	D	D	D	D
3	D	D	D	D	D	D	D
4	D	D	D	D	D	D	D
5	D	D	D	D	D	D	D
6	D	ND	D	D	D	D	D
7	D	ND	D	D	D	D	D
8	D	ND	D	D	D	D	D
9	D	ND	D	D	D	D	D
10	D	ND	D	D	D	D	D
11	D	ND	D	D	D	D	D
12	D	ND	ND	ND	ND	D	D
13	D	ND	ND	ND	ND	ND	ND
14	D	ND	ND	ND	ND	ND	ND
15	D	ND	ND	ND	ND	ND	ND
16	D	ND	ND	ND	ND	ND	ND
17	D	ND	ND	ND	ND	ND	ND
18	D	ND	ND	ND	ND	ND	ND
19	D	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND

Tabulka 2.20: Výsledky policy iteration

	Iterácia						
	0	1	2	3	4	5	6
P	D	D	D	D	D	D	D
0	D	D	D	D	D	D	D
1	D	D	D	D	D	D	D
2	D	ND	D	D	D	D	D
3	D	ND	ND	ND	ND	ND	ND
4	D	ND	ND	ND	ND	ND	ND
5	D	ND	ND	ND	ND	ND	ND
6	D	ND	ND	ND	ND	ND	ND
7	D	ND	ND	ND	ND	ND	ND
8	D	ND	ND	ND	ND	ND	ND
9	D	ND	ND	ND	ND	ND	ND
10	D	ND	ND	ND	ND	ND	ND
11	D	ND	ND	ND	ND	ND	ND
12	D	ND	ND	ND	ND	ND	ND
13	D	ND	ND	ND	ND	ND	ND
14	D	ND	ND	ND	ND	ND	ND
15	D	ND	ND	ND	ND	ND	ND
16	D	ND	ND	ND	ND	ND	ND
17	D	ND	ND	ND	ND	ND	ND
18	D	ND	ND	ND	ND	ND	ND
19	D	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND

Tabulka 2.21: Výsledky Policy iteration-mesto 5

Uvedomme si, že v každej iterácii policy iteration algoritmu sa rieši sústava rovníc o 111 neznámych. Pozrime sa preto na výsledky, ktoré dá menej zaťažujúci value-iteration algoritmus. Nastavme $\epsilon = 0.001$, takže priemerný výnos výsledného riadenia

sa od optimálneho nesmie líšiť o viac ako 0.1%. Ako už bolo spomenuté vyššie, pracujeme so zápornými výnosmi. Value iteration algorimus v tvare ako sme ho odvodili však vyžaduje splnenie $c_i(r_i) > 0,\ i \in S$ pre akékoľvek rozhodnutie $r_i \in K(i)$. Ku každej hodnote c_i preto pripočítame konštantu 350. Ak algoritmus spustíme s týmto nastavením, nikdy neskonverguje. Dôvodom je, že nie je splnená podmienka neperiodicity reťazca. K odstráneniu tohto problému použijeme transformáciu dát založenú na nasledujúcej myšlienke. Vieme, že pre stacionárne riadenie $\mathcal{R}$ s využitím vektorového zápisu platí

$$\mathbf{w}(\mathcal{R}) + g(\mathcal{R})\mathbf{e} = \mathbf{c}(\mathcal{R}) + \mathbf{P}(\mathcal{R})\mathbf{w}(\mathcal{R}),$$

čo je ekvivalentné s

$$\mathbf{w}(\mathcal{R}) + \frac{g(\mathcal{R})}{2}\mathbf{e} = \frac{\mathbf{c}(\mathcal{R})}{2} + \left(\frac{\mathbf{P}(\mathcal{R}) + \mathbf{I}}{2}\right)\mathbf{w}(\mathcal{R}).$$

Ak položíme $\bar{\mathbf{P}}(\mathcal{R}) = \frac{\mathbf{P}(\mathcal{R})+\mathbf{I}}{2}$ a $\bar{\mathbf{c}}(\mathcal{R}) = \frac{\mathbf{c}(\mathcal{R})}{2}$, tak rovnica

$$\mathbf{v} + g\mathbf{e} = \bar{\mathbf{c}}(\mathcal{R}) + \bar{\mathbf{P}}(\mathcal{R})\mathbf{v},$$

má riešenie $\mathbf{v} = \bar{\mathbf{w}}(\mathcal{R})$ a $g = \bar{g}(\mathcal{R}) = \frac{g(\mathcal{R})}{2}$, a preto by sme pri vyššie zmienenej transormácii dát dospeli s využitím policy iteration algoritmu k rovnakému riešeniu. Navyše je pre akékoľvek stacionárne riadenie matica $\bar{\mathbf{P}}(\mathcal{R})$ neperiodická. Ak teda v každej iterácii value iteration algoritmu pracujeme s $\bar{\mathbf{P}}(\mathcal{R})$ s ohodnotením $\bar{\mathbf{c}}(\mathcal{R}) + 350/2$, kde $\mathcal{R}$ je vstupujúce stacionárne riadenie, tak splníme všetky predpoklady konvergencie algoritmu a obdržíme požadovaný výsledok. Zvoľme aj pri tomto nastavení $\epsilon = 0.001$. Algoritmus potom skonveguje po 30 iteráciach, pričom dostaneme rovnaký výsledok ako v prípade použitia policy iteration algoritmu. Tento value iteration algoritmus taktiež skonvergue asi o 15% rýchlejšie ako policy iteration. V nasledujúcich tabuľkách 2.22, 2.23 a 2.24 sú hodnoty výstupných riadení pre iterácie, v ktorých sa vstupné a výstupné riadenie líši a posledné 30. vygenerované riadenie.

(a) Riadenie pre mesto 1

	Iterácia					
	1	2	3	4	5	30
P	D	D	D	D	D	D
0	ND	D	D	D	D	D
1	ND	D	D	D	D	D
2	ND	D	D	D	D	D
3	ND	ND	ND	D	D	D
4	ND	ND	ND	ND	ND	ND
5	ND	ND	ND	ND	ND	ND
6	ND	ND	ND	ND	ND	ND
7	ND	ND	ND	ND	ND	ND
8	ND	ND	ND	ND	ND	ND
9	ND	ND	ND	ND	ND	ND
10	ND	ND	ND	ND	ND	ND
11	ND	ND	ND	ND	ND	ND
12	ND	ND	ND	ND	ND	ND
13	ND	ND	ND	ND	ND	ND
14	ND	ND	ND	ND	ND	ND
15	ND	ND	ND	ND	ND	ND
16	ND	ND	ND	ND	ND	ND
17	ND	ND	ND	ND	ND	ND
18	ND	ND	ND	ND	ND	ND
19	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND

(b) Riadenie pre mesto 2

	Iterácia							
	1	2	3	4	5	6	10	30
P	D	D	D	D	D	D	D	D
0	ND	D	D	D	D	D	D	D
1	ND	D	D	D	D	D	D	D
2	ND	D	D	D	D	D	D	D
3	ND	D	D	D	D	D	D	D
4	ND	D	D	D	D	D	D	D
5	ND	ND	D	D	D	D	D	D
6	ND	ND	ND	D	D	D	D	D
7	ND	ND	ND	D	D	D	D	D
8	ND	ND	ND	ND	D	D	D	D
9	ND	ND	ND	ND	ND	D	ND	ND
10	ND	ND	ND	ND	ND	ND	ND	ND
11	ND	ND	ND	ND	ND	ND	ND	ND
12	ND	ND	ND	ND	ND	ND	ND	ND
13	ND	ND	ND	ND	ND	ND	ND	ND
14	ND	ND	ND	ND	ND	ND	ND	ND
15	ND	ND	ND	ND	ND	ND	ND	ND
16	ND	ND	ND	ND	ND	ND	ND	ND
17	ND	ND	ND	ND	ND	ND	ND	ND
18	ND	ND	ND	ND	ND	ND	ND	ND
19	ND	ND	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND	ND

Tabulka 2.22: Výsledky value iteration

(a) Riadenie pre mesto 3

	Iterácia					
	1	2	3	5	11	30
P	D	D	D	D	D	D
0	ND	D	D	D	D	D
1	ND	D	D	D	D	D
2	ND	D	D	D	D	D
3	ND	D	D	D	D	D
4	ND	ND	D	D	D	D
5	ND	ND	ND	D	ND	ND
6	ND	ND	ND	ND	ND	ND
7	ND	ND	ND	ND	ND	ND
8	ND	ND	ND	ND	ND	ND
9	ND	ND	ND	ND	ND	ND
10	ND	ND	ND	ND	ND	ND
11	ND	ND	ND	ND	ND	ND
12	ND	ND	ND	ND	ND	ND
13	ND	ND	ND	ND	ND	ND
14	ND	ND	ND	ND	ND	ND
15	ND	ND	ND	ND	ND	ND
16	ND	ND	ND	ND	ND	ND
17	ND	ND	ND	ND	ND	ND
18	ND	ND	ND	ND	ND	ND
19	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND

(b) Riadenie pre mesto 5

	Iterácia				
	1	2	3	4	30
P	D	D	D	D	D
0	ND	D	D	D	D
1	ND	ND	D	D	D
2	ND	ND	ND	D	D
3	ND	ND	ND	ND	ND
4	ND	ND	ND	ND	ND
5	ND	ND	ND	ND	ND
6	ND	ND	ND	ND	ND
7	ND	ND	ND	ND	ND
8	ND	ND	ND	ND	ND
9	ND	ND	ND	ND	ND
10	ND	ND	ND	ND	ND
11	ND	ND	ND	ND	ND
12	ND	ND	ND	ND	ND
13	ND	ND	ND	ND	ND
14	ND	ND	ND	ND	ND
15	ND	ND	ND	ND	ND
16	ND	ND	ND	ND	ND
17	ND	ND	ND	ND	ND
18	ND	ND	ND	ND	ND
19	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND

Tabulka 2.23: Výsledky value iteration

	Iterácia								
	1	2	3	4	5	6	7	27	30
P	D	D	D	D	D	D	D	D	D
0	ND	D	D	D	D	D	D	D	D
1	ND	D	D	D	D	D	D	D	D
2	ND	D	D	D	D	D	D	D	D
3	ND	D	D	D	D	D	D	D	D
4	ND	D	D	D	D	D	D	D	D
5	ND	D	D	D	D	D	D	D	D
6	ND	ND	D	D	D	D	D	D	D
7	ND	ND	D	D	D	D	D	D	D
8	ND	ND	ND	D	D	D	D	D	D
9	ND	ND	ND	ND	D	D	D	D	D
10	ND	ND	ND	ND	ND	D	D	D	D
11	ND	ND	ND	ND	ND	ND	D	D	D
12	ND	ND	ND	ND	ND	ND	ND	D	D
13	ND	ND	ND	ND	ND	ND	ND	ND	ND
14	ND	ND	ND	ND	ND	ND	ND	ND	ND
15	ND	ND	ND	ND	ND	ND	ND	ND	ND
16	ND	ND	ND	ND	ND	ND	ND	ND	ND
17	ND	ND	ND	ND	ND	ND	ND	ND	ND
18	ND	ND	ND	ND	ND	ND	ND	ND	ND
19	ND	ND	ND	ND	ND	ND	ND	ND	ND
20	ND	ND	ND	ND	ND	ND	ND	ND	ND

Tabulka 2.24: Výsledky value iteration pre mesto 4

2.7 Súvislosť s úlohou lineárneho programovania

V tejto kapitole ukážeme ako predchádzajúce výsledky vedú k možnosti prihliadať na problém nájdenia optimálneho riadenia ako na úlohu lineárneho programovania. Postup, ktorý tu stručne popíšeme, je rozsiahlejšie študovaný v knihe [7]. Predpoklad jedinej triedy trvalých stavov pre všetky stacionárne riadenia je v tomto možné dokonca zoslabiť. Predpokladajme, že všetky stacionárne riadenia, ktoré sú optimálne k dlhodobému priemernému výnosu majú jedinú triedu trvalých stavov. V prípade neoptimálnych riadení teda môžeme predpoklad jedninej triedy trvalých stavov vypustiť. Nech je v zmysle predešlého odstavca $\tilde{g}, \tilde{v}_i,\ i \in S$ riešenie rovníc (2.36). Potom existuje riešenie $g, v_i,\ i \in S$ sústavy nerovníc

$$v_i \geq c_i(r_i) - g + \sum_{j \in S} p_{ij}(r_i) v_j,\ r_i \in K(i),\ i \in S \tag{2.51}$$

Nie je ťažké presvedčiť sa, že tvrdenie vety 2.8 za zoslabeného predpokladu je možné zovšeobecniť na tvar, podľa ktorého z predošlej rovnice plynie vzťah $g \geq g_i(\mathcal{R})$, pre akékoľvek $i \in S$ a akékoľvek riadenie $\mathcal{R}$. To znamená, že pre akékoľvek riešenie (2.51) je $g \geq \tilde{g}$. Z predošlých úvah potom plynie, že $\tilde{g}$ je optimálna hodnota účelovej funkcie lineárnej úlohy

$$\min g \tag{2.52}$$

za podmienok

$$g + v_i - \sum_{j \in S} p_{ij}(r_i) v_j \geq c_i(r_i),\ r_i \in K(i),\ i \in S$$

$$g, v_i \in \Re.$$

Algoritmus lineárneho programovania je možné zostaviť nasledujúcim spôsobom

- Krok 1 - úloha lineárneho programovania
  Simplexovou metódou nájdeme optimálne riešenie $x^*_{ir},\ r \in K(i),\ i \in S$ nasledujúcej lineárnej úlohy:

  $$\max \sum_{i \in S} \sum_{r \in K(i)} c_i(r) x_{ir} \tag{2.53}$$

  za podmienok

  $$\sum_{r \in K(j)} x_{jr} - \sum_{i \in S} \sum_{r \in K(i)} p_{ij}(r) x_{ir} = 0,\ j \in S$$

  $$\sum_{i \in S} \sum_{r \in K(i)} x_{ir} = 1$$

  $$x_{ir} \geq 0,\ r \in K(i),\ i \in S$$

  Táto úloha je duálna k úlohe (2.52). K optimálnemu riešeniu $x^*_{ir},\ r \in K(i),\ i \in S$ teda náleží komplementárne riešenie duálnej úlohy $g^* = \tilde{g}, v^*_i,\ i \in S$.

- Krok 2 - zostrojenie optimálneho riadenia: inicializácia
  Zostrojíme neprázdnu množinu $I_0 := \{i : \sum_{r \in K(i)} x^*_{ir} > 0\}$. Pre $i \in I_0$ nastavíme

  $$R^*_i := r \text{ pre ľubovolné r, také že } x^*_{ir} > 0.$$

  Ďalej polož $I := I_0$

- Krok 3 - zostrojenie optimálneho riadenia: ukončovacie pravidlo
  Ak $I = S$ algoritmus zastavíme. Inak zvolíme $i \notin S$ a rozhodnutie $r \in K(i)$ také že $p_{ij}(r) > 0$ pre nejaké $j \in S$. Ďalej položíme $R^*_i := r$ a $I := I \cup \{i\}$ a pokračujeme krokom 3.

Výsledné riadenie splňuje $\mathcal{R}^* = \tilde{\mathcal{R}}$. Podľa [7] je totiž možné ukázať, že $I = S$, pričom I_0 pozostáva za riadenia $\mathcal{R}^*$ z trvalých stavov a $I - I_0$ z prechodných stavov. Ďalej riešenie splňuje

$$g^* + v^*_i - \sum_{j \in S} p_{ij}(R^*_i) v^*_j = c_i(R^*_i),\ i \in I_0, \tag{2.54}$$

a tak je podľa zovšeobecnenia vety 2.8 $g_i(R^*) = g^*,\ i \in I_0$. Z prechodného stavu sa v konečnom čase dostaneme do množiny stavov trvalých, preto aj pre prechodné stavy dostaneme $g_i(R^*) = g^*,\ i \in I - I_0$.

2.8 Modifikovaný algoritmus pre diskontovanie

V prípade, že je diskontný faktor β blízko 1 value iteration algoritmus z kapitoly 2.4.2 konverguje dosť pomaly. Vykonajme preto nasledujúcu transformáciu. Pripomeňme, že existujú jednoznačné čísla $\tilde{V}_i^\beta,\ i \in S$ splňajúce

$$\tilde{V}_i^\beta = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i) \tilde{V}_j^\beta \right\}, i \in S. \tag{2.55}$$

Od oboch strám naždej rovnice sústavy (2.55) odčitajme $\tilde{V}_N^\beta$. Po menšej úprave dostaneme

$$\tilde{V}_i^\beta - \tilde{V}_N^\beta = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i)(\tilde{V}_j^\beta - \tilde{V}_N^\beta) - (1-\beta)\tilde{V}_N^\beta \right\}, i \in S.$$

Položíme $\tilde{h}_i = \tilde{V}_i^\beta - \tilde{V}_N^\beta,\ i \in S$ a predošlú sústavu upravíme na tvar

$$\tilde{h}_i + (1-\beta)\tilde{V}_N^\beta = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \beta \sum_{j \in S} p_{ij}(r_i) \tilde{h}_j \right\}, i \in S.$$

Keďže je $\tilde{h}_N = 0$ môžeme ďalej písať

$$\tilde{h}_i + (1-\beta)\tilde{V}_N^\beta = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} \bar{p}_{ij}(r_i) \tilde{h}_j \right\}, i \in S.$$

kde sme položili $\bar{p}_{ij}(r_i) = \beta p_{ij}(r_i),\ j < N$ a $\bar{p}_{iN}(r_i) = \beta p_{iN}(r_i) + (1-\beta)$. Matica $\bar{\mathbf{P}}(\mathcal{R}) = (\bar{p}_{ij}(r_i))_{i,j \in S}$ je potom stochastická matica. Transformácia nám umožnuje iteračne počítať hodnoty

$$H_i(n+1) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} \bar{p}_{ij}(r_i) H_j(n) \right\}, \tag{2.56}$$

pričom z VI pre priemerný výnos plynie, že $\forall i \in S$ platí

$$H_i(n+1) - H_i(n) \to (1-\beta)\tilde{V}_N^\beta$$
$$H_i(n) - H_j(n) \to \tilde{h}_i - \tilde{h}_j = \tilde{V}_i^\beta - \tilde{V}_j^\beta$$

Vieme,že hodnoty $H_i(n)$ rastú do nekonečna a preto vykonajme korekciu podľa poznámky 2.20. Položíme teda $h_i(n) = H_i(n) - H_N(n),\ i \in S$ a $k_N(n+1) = H_N(n+1) - H_N(n)$, čím dostaneme

$$h_i(n+1) + k_N(n+1) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} \bar{p}_{ij}(r_i) h_j(n) \right\}, \tag{2.57}$$

kde $\forall n \in N$ je $h_N(n) = 0$.

Položme

$$L_N(n) = 1/(1-\beta) \min_{i \in S}[h_i(n+1) + k_N(n+1) - h_i(n)] \tag{2.58}$$

$$U_N(n) = 1/(1-\beta) \max_{i \in S}[h_i(n+1) + k_N(n+1) - h_i(n)] \tag{2.59}$$

$$L_i(n) = L_N(N) + h_k(n), \quad U_i(n) = U_N(n) + h_k(n),\ i \neq N \tag{2.60}$$

Potom môžeme pre hľadanie optimálneho riadenia použiť nasledovný modifikovaný value iteration algoritmus.

- Krok 0 - Inicializácia
  Položíme $h_i(0) = 0,\ i \in S$. Ďalej nastavíme $n := 1$ a postúpime na samotný iteračný algoritmus.

- Krok 1 - Value iteration
  Spočítame hodnotu

$$k_N(n) = \max_{r_N \in K(N)} \left\{ c_N(r_N) + \sum_{j \in S} \bar{p}_{Nj}(r_N) h_j(n-1) \right\}, \tag{2.61}$$

  Pre každý stav $i \in S$ dopočítame

$$h_i(n) = \max_{r_i \in K(i)} \left\{ c_i(r_i) + \sum_{j \in S} \bar{p}_{ij}(r_i) h_j(n-1) - k_N(n) \right\}.$$

  Nech $\mathcal{R}$ je stacionárne riadenie ktorého príslušné rozhodnutia $R_i,\ i \in S$ maximalizujú pravú stranu výrazov.

  Poznámka 2.21 *Na tomto mieste je vhodné upozorniť, že pri výpočte* $\mathbf{c}(\mathcal{R})$ *musíme použiť pôvodnú maticu* $\mathbf{P}(\mathcal{R})$.

- Krok 2 - Hranice
  Spočítame hranice

$$L_N(n) = 1/(1-\beta) \min_{i \in S}[h_i(n+1) - k_N(n+1) - h_i(n)] \tag{2.62}$$

$$U_N(n) = 1/(1-\beta) \max_{i \in S}[h_i(n+1) - k_N(n+1) - h_i(n)] \tag{2.63}$$

$$L_i(n) = L_N(N) + h_k(n), \quad U_i(n) = U_N(n) + h_k(n),\ i \neq N \tag{2.64}$$

- Krok 3 - Ukončovacie pravidlo

  Ak $0 \leq U_i(n) - L_i(n) \leq \epsilon$ pre $\forall i \in S$ kde $\epsilon > 0$ je predom určený koeficient presnosti, algoritmus skončí a za optimálne riadenie vezmeme $\mathcal{R}$. Volíme napríklad $\epsilon = 10^{-3}$. V opačnom prípade položíme $n := n + 1$ a pokračujeme krokom 1.

Vráťme sa k úlohe o taxikárovi a vyriešme ju pomocou modifikovaného algoritmu. Pripomeňme, že pôvodný algoritmus skonvergoval po 63 iteráciach. Modifikovaný algoritmus skonverguje s nastavením $\epsilon = 0.001$ už po 5 iteráciách. Výsledky algoritmu sú rozpísané v nasledujúcich tabuľkách.

It. (n)	Vypočítané hodnoty							Riadenie
	$h_1(n)$	$h_2(n)$	$h_3(n)$	$h_4(n)$	$h_5(n)$	$h_6(n)$	$h_7(n)$	
1	2.48	-3.15	-5.23	-3.10	-0.93	0.48	0.00	$\{2,2,2,3,4,2,4\}$
2	2.47	-3.33	-5.35	-2.76	-0.41	0.90	0.00	$\{2,2,3,1,4,2,4\}$
3	2.33	-3.47	-5.47	-2.83	-0.58	0.81	0.00	$\{2,2,3,1,4,2,4\}$
4	2.34	-3.46	-5.45	-2.81	-0.54	0.83	0.00	$\{2,2,3,1,4,2,4\}$
5	2.33	-3.47	-5.46	-2.82	-0.55	0.82	0.00	$\{2,2,3,1,4,2,4\}$

Tabulka 2.25: Riešenie - modifikované value iteration

It. (n)	Hranice							
	$L_1(n)$	$U_1(n)$	$L_2(n)$	$U_2(n)$	$L_3(n)$	$U_3(n)$	$L_4(n)$	$U_4(n)$
1	147.975	224.975	142.350	219.350	140.275	217.275	142.400	219.400
2	183.229	190.097	177.431	184.300	175.408	182.276	177.999	184.868
3	185.718	187.349	179.914	181.544	177.917	179.547	180.556	182.186
4	186.243	186.655	180.443	180.855	178.455	178.867	181.094	181.506
5	186.335	186.465	180.535	180.665	178.542	178.672	181.183	181.312

Tabulka 2.26: Riešenie - modifikované value iteration

It. (n)	Hranice					
	$L_5(n)$	$U_5(n)$	$L_6(n)$	$U_6(n)$	$L_7(n)$	$U_7(n)$
1	144.575	221.575	145.975	222.975	145.500	222.500
2	180.343	187.211	181.654	188.522	180.757	187.625
3	182.810	184.440	184.198	185.828	183.386	185.017
4	183.369	183.781	184.734	185.146	183.904	184.316
5	183.454	183.584	184.825	184.955	184.003	184.132

Tabulka 2.27: Riešenie - modifikované value iteration

Kapitola 3

Riadenie spojitých reťazcov

3.1 Úvod

V tejto kapitole definujeme riadený spojitý Markovov proces s ohodnotením, pričom využijeme výsledky z kapitoly 1.3. Zmienený proces môžeme popísať nasledujúcimi bodmi.

- Uvažujeme dynamický systém, ktorý je tentokrát sledovaný nepretržite. Systém sa môže nachádzať v jednom z N stavov. Opäť značíme $S = \{1, \ldots, N\}$ množinu prípustných stavov reťazca.

- Kontrolórovi je umožnené riadiť systém v akomkoľvek čase $t \in [0, \infty)$, pričom pre každý stav $i \in S$ má k dispozícii konečnú množinu možných akcii, ktoré môže vykonať, značenú opäť $K(i)$.

- Kým v diskrétnom čase sme riadením reťazca vo všeobecnosti rozumeli postupnosť rozhodnutí v každom časovom bode, v spojitom prípade pôjde o zprava spojitú po častiach konštantnú funkciu času. Je teda $\mathcal{R} = \mathbf{R}(t)$, kde $\mathbf{R}(t) = (R_1(t), R_2(t), \ldots, R_N(t))'$. Tento predpoklad vlastne znamená, že okamihy v ktorých kontrolór nahliadne do systému a má možnosť zmeniť jeho riadenie, tvoria diskrétnu rastúcu postupnosť. Ak nebude povedané inak, obmedzíme sa v ďalšom texte na stacionárne riadenia, t.j. $\mathcal{R} = \mathbf{R}(t) = \mathbf{R}$ je v čase konštantný vektor.

- Pre akékoľvek pevne zvolené riadenie $\mathcal{R}$ správanie systému popisuje náhodný proces $X^c(\mathcal{R}) = \{X_t(\mathcal{R}), t \geq 0\}$, ktorý je Markovským procesom so spojitým časom. Všeobecne je tento reťazec nehomogénny. S predchádzajúcich požiadaviek na systém je totiž v každom konštantnom úseku riadenia použitá iná matica intenzít prechodu. Avšak v prípade stacionárneho riadenia je rovnako ako v diskrétnom prípade reťazec homogénny, t.j. dynamika systému je popísaná jedinou maticou intenzít prechodov, ktorú označíme ako $\mathbf{Q}(\mathcal{R})$.

- Uvedomme si, že kontrolór svojim rozhodnutím vlastne určí skladbu matice intezity prechodov až do prípadnej zmeny, ktorú môže vykonať v budúcnosti. Budeme značiť $q_{ij}(R_i(t)), R_i(t) \in K(i)$ intezitu prechodu zo stavu i do stavu j za predpokladu, že v čase t vykoná kontrolór rozhodnutie $R_i(t)$. V prípade stacionárneho riadenia vynecháme argument času, pretože z vyššie uvedeného je zrejmé, že nemôže dôjsť k zmene rozhodnutia v konkrétnom stave v čase.

- Systém ohodnotenia zavedieme v zmysle kapitoly 1.3. Prechod medzi stavom i a j realizuje zisk resp. náklad z_{ij}. Taktiež za každú časovú jednotku, po ktorú reťazec zotrvá v stave $i \in S$ obdržíme resp. zaplatíme čiastku z_i.

Pracujme na množine stacionárnych riadení. Pre stacionárne riadenie $\mathcal{R}$ značíme $V_i(T, \mathcal{R}) =$ očakávaný výnos do času T ak je reťazec riadený pravidlom $\mathcal{R}$ a východiskový stav je i. V zmysle (1.18) a (1.19) potom píšeme

$$V_i(T, \mathcal{R}) = \int_0^T \sum_{j \in S} p_{ij}(t, \mathcal{R}) c_j(R_j), \tag{3.1}$$

kde

$$c_i(R_i) = z_i + \sum_{j \in S, j \neq i} q_{ij}(R_i) z_{ij},\ i \in S. \tag{3.2}$$

Všimnime si zjavnú analógiu vzorca (3.1) s diskrétnym stavom, t.j. so vzorcom (2.5). Zisky z_i a $z_{ij},\ i, j \in S$ môžeme uvažovať závislé na rozhodnutí v stave i. Potom by sme písali

$$c_i(R_i) = z_i(R_i) + \sum_{j \in S, j \neq i} q_{ij}(R_i) z_{ij}(R_i),\ i \in S.$$

Poznámka 3.1 *Pripomeňme, že v prípade riadenia diskrétneho reťazca, $c_i(R_i)$ udáva očakávaný výnos za jedno časové obdobie, t.j. jeden prechod pri výstupe zo stavu i (respektíve cenu, ktorú musíme v stave i zaplatiť aby sme mohli použiť riadenie R_i). V spojitom prípade sme v kapitole 1.3 ukázali, že $\frac{z_i}{q_i(R_i)} + \sum_{j \in S, j \neq i} \frac{q_{ij}(R_i)}{q_i(R_i)} z_{ij}$ udáva očakávaný výnos do prvého výstupu (vrátane) z počiatočného stavu i pri použití R_i. Stačí pripomenúť, že $\frac{1}{q_i(R_i)}$ je očakávaná doba zotrvania v stave i pri použití R_i a $\frac{q_{ij}(R_i)}{q_i(R_i)}$ je pravdepodobnosť, že reťazec vystúpi práve do stavu j pri použití R_i. Vzorec (3.2) potom dostaneme vydelením očakávaného výnosu do prvého výstupu z i pri použití R_i strednou dobou zotrvania v stave i pri použití R_i.*

Ak zahrnieme diskontný faktor $\beta > 0$ budeme podľa (1.20) písať

$$V_i^\beta(T, \mathcal{R}) = \int_0^T \sum_{j \in S} e^{-\beta t} p_{ij}(t, \mathcal{R}) c_j(R_j) dt \tag{3.3}$$

Pre celkový diskontovaný výnos definovaný ako

$$V_i^\beta(\mathcal{R}) = \lim_{T \to \infty} V_i^\beta(T, \mathcal{R}) \tag{3.4}$$

použijeme výsledok (1.24) a budeme písať

$$\mathbf{V}^{\beta}(\mathcal{R}) = \lim_{T\to\infty} \mathbf{V}^{\beta}(T,\mathcal{R}) = \beta^{-1}[\mathbf{c}(\mathcal{R}) + \mathbf{Q}(\mathcal{R})\mathbf{V}^{\beta}(\mathcal{R})]$$

Ďalej podľa (1.22) definujeme priemerný očakávaný výnos za časovú jednotku pri počiatočnom stave i a použití riadenia $\mathcal{R}$ ako

$$g_i(\mathcal{R}) = \lim_{T\to\infty} \frac{V_i(T,\mathcal{R})}{T}. \tag{3.5}$$

Ak má pre stacionárne riadenie $\mathcal{R}$ vnorený reťazec len jednu triedu trvalých stavov tak existuje jednoznačné stacionárne rozdelenie $\{\Pi_j(\mathcal{R}),\ j \in S\}$, pre ktoré je $\Pi_j(\mathcal{R}) = \lim_{T\to\infty} \mathbf{p}_{ij}(T,\mathcal{R}),\ i,j \in S$. Podľa (1.25) priemerný výnos nezávisí na počiatočnom stave a môžeme písať

$$g(\mathcal{R}) = \sum_{j\in S} c_j(R_j)\Pi_j(\mathcal{R}), \tag{3.6}$$

kde

$$\sum_{i\in S} \Pi_i(\mathcal{R}) q_{ij}(R_i) = 0 \tag{3.7}$$

3.2 Riadenie spojitých reťazcov

V tejto kapitole odvodíme policy iteration algortmy pre hľadanie oprimálneho riadenia vzhľadom k priemernému očakávanému výnosu a celkovému diskontovanému výnosu pričom použijeme analogoickú konštrukciu ako v prípade diskrétnych reťazcov. Pozrime sa najprv na prípad keď je kritériom diskontovaný výnos.

Veta 3.2 *Nech pre dané čísla $v_i^{\beta},\ i \in S$ platí*

$$c_i(R_i) + \sum_{j\in S} q_{ij}(R_i) v_j^{\beta} \geq \beta v_i^{\beta},\ \forall i \in S.$$

Potom pre dlhodobý diskontovaný výnos platí

$$V_i^{\beta}(\mathcal{R}) \geq v_i^{\beta},\ i \in S.$$

Dôkaz Pre $\forall i \in S$ platí

$$\begin{aligned} V_i^{\beta}(\mathcal{R}) - v_i^{\beta} &\geq \beta^{-1} c_i(R_i) + \beta^{-1} \sum_{j\in S} q_{ij}(R_i) V_j^{\beta}(\mathcal{R}) - \beta^{-1} \left(c_i(R_i) + \sum_{j\in S} q_{ij}(R_i) v_j^{\beta} \right) \\ &\geq \beta^{-1} \sum_{j\in S} q_{ij}(R_i)(V_j^{\beta}(\mathcal{R}) - v_j^{\beta}). \end{aligned}$$

Čo môžeme maticovo zapísať ako

$$\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta} \geq \beta^{-1}\mathbf{Q}(\mathcal{R})(\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}),$$

alebo ekvivalentne ako

$$(\beta\mathbf{I} - \mathbf{Q}(\mathcal{R}))(\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}) \geq 0$$

Ďalej nech $B > 0$. Vynásobením oboch strán merovnosti hodnotou B^{-1} a pričítaním $(\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}(\mathcal{R}))$ nakoniec dostaneme

$$(\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}) \geq [B^{-1}(\mathbf{Q}(\mathcal{R}) - \beta\mathbf{I}) + \mathbf{I}](\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}).$$

Ak označíme $\bar{P}(\mathcal{R}) = [B^{-1}(\mathbf{Q}(\mathcal{R}) - \beta\mathbf{I}) + \mathbf{I}]$, tak pre B dostatočne veľké má táto matica všetky prvky kladné, menšie ako 1 so spektrálnym polomerom menším ako 1. Je teda $(\mathbf{I} - \bar{\mathbf{P}}(\mathcal{R}))(\mathbf{V}^{\beta}(\mathcal{R}) - \mathbf{v}^{\beta}) \geq \mathbf{0}$ pričom je $(\mathbf{I} - \bar{\mathbf{P}}(\mathcal{R}))^{-1} = \sum_{k=0}^{\infty} \bar{\mathbf{P}}^{k} \geq \mathbf{0}$ □

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $v_i^{\beta}(\mathcal{R}),\ i \in S$ sústavy lineárnych rovníc

  $$\beta v_i^{\beta} \quad = \quad c_i(R_i) + \sum_{j \in S} q_{ij}(R_i) v_j^{\beta},\ i \in S$$

  o neznámych $v_i^{\beta},\ i \in S$.

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

  $$c_i(r_i) + \sum_{j \in S} q_{ij}(r_i) v_j^{\beta}(\mathcal{R}),\ i \in S. \tag{3.8}$$

  Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizuje výraz (3.8), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S$.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1 pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$.

Veta 3.3 *Nech je dané stacionárne riadenie $\mathcal{R}$, pre ktoré má príslušný vnorený Markovov reťazec len jedinú triedu trvalých stavov. Nech pre dané čísla g a $v_i,\ i \in S$ platí*

$$c_i(R_i) - g + \sum_{j \in S} q_{ij}(R_i) v_j \geq 0,\ \forall i \in S. \tag{3.9}$$

Potom dlhodobý priemerný výnos spĺňa

$$g(\mathcal{R}) \geq g. \tag{3.10}$$

Ostrá nerovnosť bude v (3.10) *práve vtedy ak za riadenia* $\mathcal{R}$ *existuje trvalý stav* l *(v zmysle vnoreného reťazca), taký že v* (3.9) *bude pre* $i = l$ *ostrá nerovnosť.*

Dôkaz: Postupujme analogicky ako v diskrétnom prípade vo vete 2.8. Vnorený reťazec má jedinú triedu trvalých stavov, a tak máme k dispozícii jednoznačné stacionárne rozdelenie $\{\Pi_j(\mathcal{R}),\ j \in S\}$. Prenásobením rovnice (3.9) pre každé i výrazom $\Pi_i(\mathcal{R})$ a následným sčítaním cez i dostaneme

$$\sum_{i\in S} \Pi_i(\mathcal{R}) c_i(R_i) - g + \sum_{i\in S} \Pi_i(\mathcal{R}) \sum_{j\in S} q_{ij}(R_i) v_j \geq 0$$

pričom ostrú nerovnosť dostaneme práve vtedy ak máme v (3.9) ostrú nerovnosť pre nejaký stav l, taký že $\Pi_r(\mathcal{R}) > 0$, čo znamená, že l musí byť trvalým stavom vnoreného reťazca. S využitím (3.6) a (3.7) dostávame tvrdenie vety.

Definícia 3.4 Nech je dané stacionárne riadenie $\mathcal{R}$, pre ktoré má príslušný vnorený reťazec len jedinú triedu trvalých stavov. Potom definujeme relatívne hodnoty

$$w_i(\mathcal{R}) = K_i(\mathcal{R}) - g(\mathcal{R}) T_i(\mathcal{R}),\ i \in S, \tag{3.11}$$

pričom zaveďme konvenciu, že za referenčný stav l zvoľme najväčší trvalý stav v zmysle očíslovania v rámci S. Pripomeňme, že pracujeme s $S = \{1, \ldots, N\}$.

Veta 3.5 *Ak pracujeme s konečnou množinou stavov* S*, tak sú pravdepodobnosti prechodu* $p_{ij}(t)$ *diferencovatelné pre* $\forall i, j \in S$ *a* $t > 0$ *a platí*

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t),$$

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}.$$

Dôkaz: v [3] veta 3.9

Veta 3.6 *Nech je dané stacionárne riadenie* $\mathcal{R}$*, pre ktoré má príslušný vnorený Markovov reťazec len jedinú triedu trvalých stavov. Priemerný výnos* $g(\mathcal{R})$ *a relatívne hodnoty* $w_j(\mathcal{R}),\ j \in S$ *sú potom riešením sústavy lineárnych rovníc*

$$0 = c_i(R_i) - g + \sum_{j\in S} q_{ij}(R_i) v_j,\ i \in S \tag{3.12}$$

o neznámych g *a* $v_j,\ j \in S$*.*

Pre túto sústavu platí, že ak sú čísla g a $v_j,\ j \in S$ jej riešením, tak pre nejakú konštantu c platí

$$g = g(\mathcal{R}),\ v_j = w_j(\mathcal{R}) + c,\ j \in S.$$

Pre ľubovolne zvolený stav s má sústava spolu s podmienkou $v_s = 0$ jednoznačné riešenie.

Dôkaz: Pri dokazovaní tohto tvrdenia využijeme myšlienky z dôkazu vety 2.10, čím budeme demonštrovať analógiu medzi riadením spojitého a diskrétneho reťazca. Zvoľme ľubovolné $i \in S$ a dosaďme $g(\mathcal{R})$ a $w_j(\mathcal{R}), j \in S$ do i-tej rovnice sústavy (3.12). S využitím definície relatívnej hodnoty dostaneme

$$c_i(R_i) - g(\mathcal{R}) + \sum_{j \neq l,i} q_{ij}(R_i)[K_i(\mathcal{R}) - g(\mathcal{R})T_i(\mathcal{R})] - q_i w_i(\mathcal{R}).$$

Ďalej upravme na

$$c_i(R_i) + \sum_{j \neq l,i} q_{ij}(R_i)K_i(\mathcal{R}) - g(\mathcal{R}) \left[1 + \sum_{j \neq l,i} q_{ij}(R_i)T_i(\mathcal{R})\right] - q_i w_i(\mathcal{R})$$

a vyberme q_i pred zátvorku čím obdržíme

$$q_i \left\{ c_i(R_i)\frac{1}{q_i} + \sum_{j \neq l,i} \frac{q_{ij}}{q_i}(R_i)K_i(\mathcal{R}) - g(\mathcal{R}) \left[\frac{1}{q_i} + \sum_{j \neq l,i} \frac{q_{ij}}{q_i}(R_i)T_i(\mathcal{R})\right] - w_i(\mathcal{R}) \right\}.$$

Keďže pre očakávaný kumulovaný výnos a čas prvého vstupu do stavu l platí

$$\begin{aligned} T_i(\mathcal{R}) &= \frac{1}{q_i} + \sum_{j \neq l,i} \frac{q_{ij}}{q_i} T_j(\mathcal{R}),\ i \in S \\ K_i(\mathcal{R}) &= c_i(R_i)\frac{1}{q_i} + \sum_{j \neq l,i} \frac{q_{ij}}{q_i} K_j(\mathcal{R}),\ i \in S, \end{aligned}$$

dostaneme, po dosadení do rovnice 0. Tým je táto časť dôkazu dokončená.

Nech g a $v_j, j \in S$ sú ľubovolné riešenia sústavy (3.12). Už intuitívne môžeme vytušiť, že vzorec (2.8) má v prípade spojitého reťazca tvar

$$v_i = V_i(T, \mathcal{R}) - Tg + \sum_{j \in S} p_{ij}(T, \mathcal{R})v_j, i \in S \tag{3.13}$$

Za jeho platnosti, potom už ľahko dokážeme, že $g = g(\mathcal{R})$. Stačí si uvedomiť, že podobne ako v diskrétnom prípade je

$$\lim_{T \to \infty} V_i(T, \mathcal{R})/T = g(\mathcal{R})$$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{j\in S}p_{ij}(T,\mathcal{R})v_j=0,$$

a tak stačí využiť limitný prechod pre $T\to\infty$.

Platnosť vzora (3.13) môžeme dokázať s použitím kolmogorových diferenciálnych rovníc nasledujúcim spôsobom, v ktorom využijeme vektorový zápis.

$$\begin{aligned}
\mathbf{V}(T,\mathcal{R}) &= \int_0^T \mathbf{P}(t,\mathcal{R})\mathbf{c}(\mathcal{R})dt \\
&= \int_0^T \mathbf{P}(t,\mathcal{R})[\mathbf{c}(\mathcal{R})+\mathbf{Q}(\mathcal{R})\mathbf{v}]dt - \int_0^T \mathbf{P}(t,\mathcal{R})\mathbf{Q}(\mathcal{R})\mathbf{v}dt \\
&= \int_0^T \mathbf{P}(t,\mathcal{R})[\mathbf{c}(\mathcal{R})+\mathbf{Q}(\mathcal{R})\mathbf{v}]dt + [\mathbf{I}-\mathbf{P}(t,\mathcal{R})]\mathbf{v} \\
&= \int_0^T \mathbf{P}(t,\mathcal{R})g\mathbf{e}dt + [\mathbf{I}-\mathbf{P}(t,\mathcal{R})]\mathbf{v} \\
&= Tg\mathbf{e}+\mathbf{v}-\mathbf{P}(T,\mathcal{R})\mathbf{v}
\end{aligned}$$

Ďalej dokazujeme, že pre nejakú konštantu c platí $v_i = w_i(\mathcal{R})+c,\ i\in S$. Zas postupujeme analogicky ako v dôkaze vety 2.10. Nech g' a $v_i',\ i\in S$ je nejaké iné riešenie sústavy. Opäť je teda $g=g'=g(\mathcal{R})$, a tak z (3.13) dostaneme

$$v_i - v_i' = \sum_{j\in S} p_{ij}(t,\mathcal{R})[v_j - v_j'],\ i\in S,\ t>0.$$

Tentokrát zintegrujeme od 0 po T pre akékoľvek $T>0$ následne vydelíme T čím dostaneme

$$\frac{1}{T}\int_0^T [v_i - v_i']dt = \frac{1}{T}\int_0^T \sum_{j\in S} p_{ij}(t,\mathcal{R})[v_j - v_j']dt,\ i\in S,\ T>0.$$

Upravíme a zameníme poradie konešnej sumy a integrálu.

$$v_i - v_i' = \sum_{j\in S}\left(\frac{1}{T}\int_0^T p_{ij}(t,\mathcal{R})dt\right)[v_j - v_j'],\ i\in S, T>0$$

Limitným prechodom pre $T\to\infty$ obdržíme

$$v_i - v_i' = \sum_{j\in S}\Pi_j(\mathcal{R})[v_j - v_j'],\ i\in S.$$

Dospejeme k rovnakému výsledku ako v diskrétnom prípade. Tým je tvrdenie dokázané.

Pre ľubovolnú konštantu c je $v_i = w_i(\mathcal{R}) + c,\ i \in S$ riešením sústavy. Pre každé $i \in S$ je

$$\begin{aligned} 0 &= c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} q_{ij}(R_i) w_j(\mathcal{R}) \\ &= c_i(R_i) - g(\mathcal{R}) + \sum_{j \in S} q_{ij}(R_i)(w_j(\mathcal{R}) + c). \end{aligned}$$

Stačí len odčítať konštantu c z oboch strán čím dostaneme platnú rovnosť. Podmienka $v_s = 0$ potom jednoznačne určí konštantu $c = -w_s(\mathcal{R})$. Ukázali sme teda, že v tomto prípade existuje jednoznačné riešenie. □

Policy iteration algoritmus:

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $g(\mathcal{R}),\ v_i(\mathcal{R})$ sústavy lineárnych rovníc:

$$\begin{aligned} 0 &= c_i(R_i) - g + \sum_{j \in S} q_{ij}(\mathcal{R}) v_j,\ i \in S, \\ v_s &= 0, \end{aligned}$$

  kde s je ľubovolne zvolený stav. Korektnosť kroku zaručuje veta 3.6.

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

$$c_i(r_i) - g(\mathcal{R}) + \sum_{j \in S} q_{ij}(r_i) v_j(\mathcal{R}). \tag{3.14}$$

  Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizije výraz (3.14), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S$. Stačí si uvedomiť, že maximum výrazu (3.14) je pre $\forall i \in S$ väššie alebo rovné 0 a použiť vetu 3.3, podľa ktorej je $g(\mathcal{R}) \leq g(\overline{\mathcal{R}})$.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1 pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$.

Veta 3.7 *Nech $\mathcal{R}$ a $\overline{\mathcal{R}}$ sú stacionárne riadenie vygenerované algoritmom, ktoré sú bezprostredne za sebou, také že $\mathcal{R} \neq \overline{\mathcal{R}}$. Potom platí*

i) $g(\mathcal{R}) < g(\overline{\mathcal{R}})$ *alebo*
ii) $g(\mathcal{R}) = g(\overline{\mathcal{R}})$ *a* $w_i(\mathcal{R}) \leq w_i(\overline{\mathcal{R}})$ *pre* $\forall i \in S$*, pričom je nerovnosť ostrá aspoň pre jeden stav* i*.*

Dôkaz: Hlavnou myšlienkou dôkazu je prechod k vnorenému reťazcu, ktorý má diskrétny čas a tak na neho môžeme aplikovať výsledky z kapitoly 2.4.1. V prípade spojitého reťazca máme podľa kroku 2 Policy itereation algoritmu zaručenú platnosť nerovnosti

$$c_i(\overline{\mathcal{R}}) - g(\mathcal{R}) + \sum_{j \in S} q_{ij}(\overline{\mathcal{R}}) w_j(\mathcal{R}) \geq 0,$$

a teda je $g(\mathcal{R}) \leq g(\overline{\mathcal{R}})$. Za predpokladu $g(\mathcal{R}) = g(\overline{\mathcal{R}})$ dostávame $w_i(\mathcal{R}) = w_i(\overline{\mathcal{R}}),\ i \in I(\overline{\mathcal{R}})$. Stačí použiť vetu 3.3 a potom postupovať úplne analogicky ako vo vete 2.13. Ďalej rovnakým spôsobom ako vo vete 2.11 odvodíme vzťah

$$0 = \Gamma_i(\mathcal{R}, \overline{\mathcal{R}}) + \sum_{j \in S} q_{ij}(\overline{R}_i)[w_j(\overline{\mathcal{R}}) - w_j(\mathcal{R})], i \in S, \tag{3.15}$$

kde

$$\Gamma_i(\mathcal{R}, \overline{\mathcal{R}}) = c_i(\overline{R}_i) - c_i(R_i) + \sum_{j \in S} [q_{ij}(\overline{R}_i) - q_{ij}(R_i)] w_j(\mathcal{R}), i \in S.$$

Vydelením rovníc (3.15) pre $i \in S$ celkovou intenzitou $q_i(\overline{\mathcal{R}})$ prejdeme k pravdepodobnosťam prechodu vo vnorenom reťazci. Rovnice (3.15) teda môžeme prepísať na tvar

$$0 = \Gamma_i(\mathcal{R}, \overline{\mathcal{R}})/q_i(\overline{\mathcal{R}}) + \sum_{j \in S, j \neq i} \hat{p}_{ij}(\overline{R}_i)[w_j(\overline{\mathcal{R}}) - w_j(\mathcal{R})] - w_i(\overline{\mathcal{R}}) + w_i(\mathcal{R}), i \in S$$

Je teda

$$w_i(\overline{\mathcal{R}}) - w_i(\mathcal{R}) = \Gamma_i(\mathcal{R}, \overline{\mathcal{R}})/q_i(\overline{\mathcal{R}}) + \sum_{j \in S} \hat{p}_{ij}(\overline{R}_i)[w_j(\overline{\mathcal{R}}) - w_j(\mathcal{R})], \tag{3.16}$$

kde $\hat{p}_{ii}(\overline{R}_i) = 0$ a $\Gamma_i(\mathcal{R}, \overline{\mathcal{R}})/q_i(\overline{\mathcal{R}}) \geq 0$. Zaoberáme se prechodnými stavmi (ktoré za nami učinenými predokladmi musia byť v reťazci prítomné, inač by sme dostali spor s $\mathcal{R} \neq \overline{\mathcal{R}}$). Definujme vektory

$$\mathbf{u} = \{w_i(\overline{\mathcal{R}}) - w_i(\mathcal{R}),\ i \notin I(\overline{\mathcal{R}})\} \text{ a } \mathbf{\Gamma}(\mathcal{R}, \overline{\mathcal{R}}) = \{\Gamma_i(\mathcal{R}, \overline{\mathcal{R}})/q_i(\overline{\mathcal{R}}),\ i \notin I(\overline{\mathcal{R}})\}$$

Pre prechodné stavy rovnice (3.16) prepíšeme na vektorový tvar

$$\mathbf{u} = \mathbf{\Gamma}(\mathcal{R}, \overline{\mathcal{R}}) + \widetilde{P}(\overline{\mathcal{R}})\mathbf{u},$$

kde $\widetilde{P}(\overline{\mathcal{R}})$ je matica so spektrálnym polomerom menším ako 1, ktorá vznikne z $\hat{\mathbf{P}}(\overline{\mathcal{R}})$ vyškrtnutím trvalých stavov. Z rovnice dostávame $\mathbf{u} > \mathbf{0}$ (viď veta 2.11).

□

3.3 Riadenie semi-Markovských procesov

V tomto odstavci ešte stručne popíšeme policy-iteration algoritmus pre nájdenie optimálneho riadenia ak je krutériom priemerný výnos a pracujeme so semi-Markovskými procesmi. Opäť pracujeme na množine stacionárnych riadení, pre ktoré má vnorený reťazec len jednu triedu trvalých stavov. Pre stacionárne riadenie $\mathcal{R} = \{R_1, R_2, \ldots, R_N\}$ budeme v zmysle výsledkov z kapitoly 1.3 pre $\forall i \in S$ značiť

$$\begin{aligned} p_{ij}(R_i) &= \text{pravdepodobnosť, že pri voľbe rozhodnutia } R_i \text{ v stave } i \text{ dôjde po zotrvaní} \\ &\quad \text{v stave } i \text{ k prechodu práve do stavu } j \\ \tau_i(r) &= \text{očakávaná doba zotrvania v stave } i \text{ pri voľbe } R_i \\ \hat{c}_i(R_i) &= \text{očakavaný výnos do prvého výstupu (vrátane) z počiatočného stavu } i \\ &\quad \text{pri voľbe } R_i \end{aligned}$$

Podľa (1.26) máme $\hat{c}_i(R_i) = \tau_i(R_i)z_i + \sum_{j\in S} \hat{p}_{ij} z_{ij}$. Zostavenie policy iteration algoritmu je v tomto momente značne intuitívne. Z využitím výsledkov z vety 1.17 a lemmy 1.18 je opäť možné na základe teórie regeneratívnych procesov s ohodnotením dokázať analogickiu viet 2.10 resp. 3.6 a tak vytvoriť krok 2 policy iteration algoritmu. Zlepšenie riadenia sa dokáže analogickým postupom ako veta 2.8. Podobne to bude aj s konvergenciou algoritmu. Policy iteration algoritmus za predpokladu, že pre všetky stacionárne riadenia má vnorený reťazec len jednu triedu trvalých stavov potom môžeme vysloviť nasledovne.

- Krok 0 - Inicializácia
  Zvolíme ľubovolné stacionárne riadenie $\mathcal{R}$

- Krok 1 - ocenenie použitého riadenia
  Pre aktuálne pravidlo $\mathcal{R}$, spočítame jednoznačné riešenie $g(\mathcal{R}),\ v_i(\mathcal{R})$ sústavy lineárnych rovníc:

$$\begin{aligned} v_i &= \hat{c}_i(R_i) - g\tau_i(R_i) + \sum_{j\in S} p_{ij}(R_i)v_j,\ i \in S, \\ v_s &= 0, \end{aligned}$$

  kde s je ľubovolne zvolený stav.

- Krok 2 - zlepšenie použitého riadenia
  Pre $\forall i \in S$ nájdeme rozhodnutie $r_i \in K(i)$, ktoré maximalizuje výraz

$$\hat{c}_i(r_i) - g(\mathcal{R})\tau_i(r_i) + \sum_{j\in S} p_{ij}(r_i)v_j(\mathcal{R}). \tag{3.17}$$

Zostrojíme nové stacionárne riadenie $\overline{\mathcal{R}}$ tak, že položíme $\overline{R_i} = R_i$ ak pre pôvodné riadenie platí, že R_i maximalizije výraz (3.17), inak príslušné rozhodnutia zvolíme ako $\overline{R_i} = r_i,\ \forall i \in S$.

- Krok 3 - test konvergencie
  Ak nové riadnie $\overline{\mathcal{R}} = \mathcal{R}$ algoritmus sa zastaví. Inak prejdeme na krok 1 pričom za aktuálne riadenie berieme $\overline{\mathcal{R}}$.

Záver

V tejto práci sme sa zaoberali riadením Markovových reťazcov s ohodnotením. Po zavedení diskrétnych a spojitých reťazcov s ohodnotením a odvodení ich základných vlastností v kapitole 1, sme v kapitole 2 uviedli niekoľko algoritmických postupov pre hľadanie optimálneho riadenia v prípade diskrétnych systémov. Platnosť týchto algoritmov sme riadne dokázali, pričom sme tvrdenia formulovali tak aby sme ukázali analógiu medzi jednotlivými úlohami. Podarilo sa nám dokázať aj konvergenciu value iteration algoritmu pre priemerný výnos, čo sa ukázalo byť ako značne netriviálna úloha. Uviedli sme niekoľko ilustračných jednoduchých aplikácií, ktoré boli skúmané už Howardom vo svojej práci Dynamic Programming and Markov Processes a vyriešili sme aj menej triviálnu úlohu cestujujúceho opravára. Do prílohy sme zaradili program na riešenie vybraných problémov napísaný v jazyku Java. Nakoniec sa nám v kapitole 3 podarilo dokázať policy iteration algoritmi pre spojité precesy s dôrazom na poukázanie analógie s diskrétnymi stavmi.

Literatura

[1] Howard, R. A. (1960): *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA.

[2] Keogh, J. (2005): *JAVA bez předchozích znalostí*. CP Books, Brno.

[3] Prášková, Z., Lachout, P. (1998): *Základy náhodných procesů*. Karolinum, Praha.

[4] Ross, S.M. (1970): *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco, CA.

[5] Sladký, K. (2010): *Markov decision chains in discrete and continuos time; A unified approach*. Quantitative Methods in Economics (M. Reiff, ed.), Ekonomicka univerzita, Bratislava.

[6] Sladký, K. (2010): *Identification of optimal policies in Markov decision processes*. Kybernetika č.3, 558–570.

[7] Tijms, H. C. (2003): *A First Course in Stochastic Models*. Wiley, Chichester.

Prílohy

Na priloženom CD sa nachádza elektronická podoba práce a program napísaným v programovacom jazyku JAVA so stručným návodom na použitie. Program je nástrojom na riešenie numerických úloh pre vybrené problémy.